Rec'd PCT 08 DEC 2004

PCT/EP 03 / 06096

INVESTOR IN PEOPLE

PRIORITY DOCUMENT
SUBMITTED OR TRANSMITTED IN COMPLIANCE WITH RULE 17.1(a) OR (b)

REC'D 14 AUG 2003
WIPO PCT

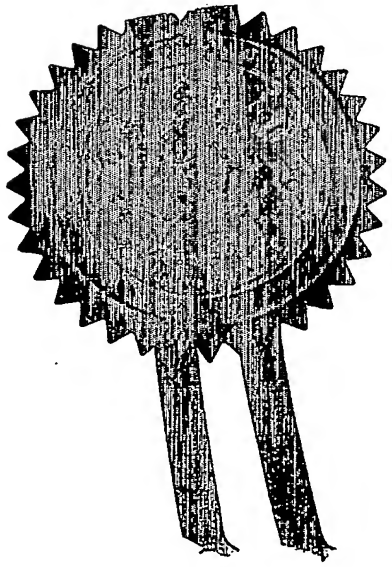
The Patent Office
Concept House
Cardiff Road
Newport
South Wales
NP10 8QQ

I, the undersigned, being an officer duly authorised in accordance with Section 74(1) and (4) of the Deregulation & Contracting Out Act 1994, to sign and issue certificates on behalf of the Comptroller-General, hereby certify that annexed hereto is a true copy of the documents as originally filed in connection with the patent application identified therein.

In accordance with the Patents (Companies Re-registration) Rules 1982, if a company named in this certificate and any accompanying documents has re-registered under the Companies Act 1980 with the same name as that with which it was registered immediately before re-registration save for the substitution as, or inclusion as, the last part of the name of the words "public limited company" or their equivalents in Welsh, references to the name of the company in this certificate and any accompanying documents shall be treated as references to the name with which it is so re-registered.

In accordance with the rules, the words "public limited company" may be replaced by p.l.c., plc, P.L.C. or PLC.

Re-registration under the Companies Act does not constitute a new legal entity but merely subjects the company to certain additional company law rules.



Signed

Dated 15 May 2003

Patents Form 1/77

Patents Act 1977
(Rule 16)

The Patent Office

1/77

THE PATENT OFFICE
A
11 JUN 2002
LONDON

Request for grant of a patent

(See the notes on the back of this form. You can also get an explanatory leaflet from the Patent Office to help you fill in this form)

The Patent Office
Cardiff Road
Newport
Gwent NP9 1RH

1. Your reference — KC/RSG/B45311

11 JUN 2002

2. Patent application number
(The Patent Office will fill in his part) — 0213365.0

12JUN02 E724998-1 C69803
P01/7700 0.00-0213365.0

3. Full name, address and postcode of the or of each applicant *(underline all surnames)*

① GlaxoSmithKline Biologicals s.a. 8101271001
Rue de l'Institut 89, B-1330 Rixensart, , Belgium

② GlaxoGroup Limited 473645004

Belgian

see continuation sheet for further applicant(s)

Patents ADP number *(if you know it)*

If the applicant is a corporate body, give the country/state of its incorporation

4. Title of the invention — Novel Compounds

5. Name of your agent *(if you have one)*

Corporate Intellectual Property

"Address for service" in the United Kingdom to which all correspondence should be sent *(including the postcode)*

GlaxoSmithKline
Corporate Intellectual Property CN925.1
980 Great West Road
BRENTFORD
Middlesex TW8 9GS

Patents ADP number *(if you know it)* 7960982003

6. If you are declaring priority from one or more earlier patent applications, give the country and the date of filing of the or each of these earlier applications and *(if you know it)* the or each application number

Country	Priority application number *(if you know it)*	Date of filing *(day / month / year)*

7. If this application is divided or otherwise derived from an earlier UK application, give the number and the filing date of the earlier application

Number of earlier application	Date of filing *(day / month / year)*

8. Is a statement of inventorship and of right to grant of a patent required in support of this request? *(Answer yes if:*

a) any applicant named in part 3 is not an inventor, or
b) there is an inventor who is named as an applicant, or
c) any named applicant is a corporate body
See note (d)

Patents Form 1/77

9. Enter the number of sheets for any of the following items you are filing with this form. Do not count copies of the same document

Continuation sheets of this form	0
Description	39 ✓
Claim(s)	2 ✓ CY
Abstract	0
Drawings	19 only

10. If you are also filing any of the following, state how many against each item.

Priority Documents	
Translations of priority documents	
Statement of inventorship and right to grant of a patent *(Patents Form 7/77)*	
Request for preliminary examination and search *(Patents Form 9/77)*	
Request for substantive examination *(Patents Form 10/77)*	
Any other documents *(please specify)*	

11. We request the ~~grant of~~ a patent on the basis of this application
Signature _____ Date 11-Jun-02
R J Easeman

12. Name and daytime telephone number of person to contact in the United Kingdom — R J Easeman 020 80474407

Warning

After an application for a Patent has beeen filed, the Comptroller of the Patent Office will consider whether publication or communication of the invention should be prohibited or restricted under Section 22 of the Patents Act 1977. You will be informed if it is necessary to prohibit or restrict your invention in this way. Furthermore, if you live in the United Kingdom, Section 23 of the Patents Act 1977 stops you from applying for a patent abroad without first getting written permission unless an application has been filed at least six weeks beforehand in the United Kingdom for a patent for the same invention and either no direction prohibiting publication or communication has been given, or any such direction has been revoked.

Notes

a) If you need help to fill in this form or you have any questions, please contact the Patent Office on 0645 500505
b) Write your answers in capital letters using black ink or you may type them.
c) If there is not enough space for all relevant details on any part of this form, please continue on a separate sheet of paper and write "see continuation sheet" in the relevant part(s). Any continuation sheet should be attached to this form.
d) If you have answered 'Yes' Patents Form 7/77 will need to be filed.
f) For details of the fee and ways to pay please contact the Patent Office.

B45311

Vaccines

The present invention relates to fusion partners which act as immunological fusion partners, as expression enhancers, and preferably to fusion partners having both functions. The invention also relates to fusion proteins containing them, to their manufacture, to their use in vaccines and to their use in medicines. In particular fusion partners are provided that contain a so-called choline binding domain, for example fusions comprising LytA from Streptococcus pneumoniae, or the pneumococcal phage CP1 lysozyme (CPL1) wherein the choline binding domain is modified to include a heterologous T-helper epitope. Such fusion partners are shown to improve the expression level of the heterologous protein attached thereto and also find particular utility when fused to poorly immunogenic proteins or peptides that are otherwise useful as vaccine antigens. More particularly, such fusion partners are useful in constructs comprising self-antigens, eg tumour specific or tissue specific antigens.

Streptococcus pneumoniae synthesises an N acetyl-L-alanine amidase, LytA, an autolysin, that specifically degrades the peptidoglycan backbone of the cell wall eventually leading to cell lysis. Its polypeptide chain has two domains. The N-terminal domain is responsible for the catalytic activity, whereas the C-terminal domain of LytA is responsible for the affinity to choline and anchorage to the cell wall. This C-terminal domain is known to bind to choline and choline analogues, and will also bind to tertiary amines such as DEAE (diethyl amino ethyl) commonly used in chromatography.

LytA is a 318 amino acid protein, and the C-terminal part comprises a tandem of six imperfect repeats of 20 or 21 amino acids and a short COOH-terminal tail. The repeats are located at the following positions:

R1: 177-191

R2: 192-212

R3: 213-234

R4: 235-254

R5: 255-275

R6: 276-298

These repeats are predicted to be in a beta-turn conformation. The C-terminus is responsible for binding choline. Likewise the C-terminus of CPL1 is responsible for binding affinity and the aromatic residues in the repeat contribute to such binding. These

proteins have been used as affinity tags to allow for rapid purification (Sanchez Puelles, Eur J Biochem. 1992, 203, 153-9).

Other proteins with a choline-binding domain have also been studied in Streptococcus pneumoniae.

One of them PspA (or Pneumococcal Surface Protein A), is a virulence factor (Yother J and Briles (1992) J Bacteriol 174(2) p 601). This protein is antigenic and immunogenic. It has a C-terminal domain consisting of 10 repeats of 20 amino acids, homologous with repeats of LytA.

CbpA (or Choline-Binding Protein A) is involved in the adherence of the pneumococcus to human cells (Rosenow et al (1997) Mol Microbiol 25 (5) p 819). It shows 10 repeats of 20 amino acids in the C-terminal domain which are almost identical to those of PspA.

LytB and LytC have a different modular organisation from the above-mentioned proteins as their choline-binding domain, made up of 15 repeats and 11 repeats respectively, is situated at the N-terminal end, not at the C-terminal end (Garcia P Mol Microbiol (1999) 31 (4) p1275 and Garcia P et al (1999) Mol Microbiol 33(1) p128). Sequence comparison shows LytB to have glucosamidase activity. LytC shows in vitro a lysozyme-type activity.

Additionally, three genes called PepA, PepB and PepC were cloned in 1995. Although their function is unknown, these genes also have a variable number of repeats homologous to those of LytA.

In their infection cycle, phages synthesise murein hydrolases facilitating their passage into the bacterium. These hydrolases have a choline-binding domain.

The muramidase CPL1 of the phage Cp-1 has been well studied. It shows 6 repeats of 20 amino acids at the C-terminus involved in the specific recognition of choline (Garica J. L. J. Virol 61 (8) p2573-80; (1987) and Garcia E Prol Natl Acad Sci (1988) p914). A comparison of the LytA and CPL1 repeats enables an initial consensus of those repeats to be made.

The murein hydrolases of phages Dp-1 (Garcia P et al (1983) J Gen Microbiol 129 (2) p489, Cpl-9 (Garcia P et al (1989) Biochem Biophys Res Commun 158(1) p 251, HB-3 Romero et al 1990 J Bacteriol 172 (9) p 5064-5070) and EJ-1 Diaz (1992) J Bacteriol 174 (17) p 5516), also show the characteristics of choline-binding domains.

This property is also shared by the lysozyme encoded by CP-1 a pneumococal phage.

WO 99/10375 describes *inter alia*, human papilloma virus proteins E6, or E7 linked to a His tag and the C-terminal portion of LytA (herein (C-LytA) and the purification of the proteins by differential affinity chromatography.

WO 99/40188 describes *inter alia* fusion proteins comprising MAGE antigens with a His tails and a C-LytA portion at the N-terminus of the molecule.

It has now been surprisingly found that fusion partners according to the present invention, when fused to a heterologous protein were capable of enhancing the immunogenicity of the heterologous proteins attached thereto. It has also been found that the expression level of the heterologous proteins attached thereto can be enhanced. The present invention accordingly provides in a preferred embodiment an improved immunological fusion partner which can also act as an expression enhancer.

Accordingly the present invention comprises a fusion molecule comprising a choline binding domain or a fragment thereof or an analogue thereof, and a heterologous promiscuous MHC Class II T-epitope, wherein said fusion partner shows a capability of acting as both an immunological fusion partner, or as an expression enhancer and preferably as both an immunological partner and expression enhancer. A promiscuous T-helper epitope is an epitope that binds to more than one MHC Class II allele, preferably more than 3 MHC Class II alleles. In particular such epitopes are capable of eliciting helper T cell response in large numbers of individuals expressing diverse MHC haplotypes.

Optionally, the fusion protein may retain its capability to bind to choline.

In one embodiment of the present invention the modified choline binding domain (fusion partner) has a capability of acting as an expression enhancer with the resulting fusion protein will be expressed at a higher yield in a host cell as compared to the unfused protein, preferably at a yield greater than about 100% (2-fold higher) or 150% or more, as measured by SDS-PAGE followed by Coomassie blue staining or silver staining, optionally followed by gel scanning. The modified choline binding domain according to the invention has also the capability of acting as an immunological partner with the resulting fusion protein with a heterologous protein will be more immunogenic in a host as compared to the unfused heterologous protein.

In another embodiment of the present invention, the modified choline binding domain has the capability to act as an immunological fusion partner, allowing an enhanced immune response to be obtained with the fusion protein as compared to the heterologous protein alone.

In a preferred embodiment, the modified choline binding domain has a dual function, having the capability to act as both an immunological fusion partner and as an expression enhancer.

In a preferred embodiment the choline binding moiety is derived from the C terminus of LytA. Preferably the C-LytA or derivatives comprises at least four repeats. In this context, C-LytA derivatives refer to a variant of C-LytA according to the present invention, that is to say variants which have retained both the capability of acting as an immunological partner and an expression enhancer. Preferred variants include, for example, peptides comprising an amino acid sequence having at least 85% identity, preferably at least 90% identity, more preferably at least 95% identity, most preferably at least 97-99% identity, to any of the repeats R1 to R6 set forth in figure 1A (SEQ ID NO:1 to 6), or a peptide comprising an amino acid sequence having at least 15, 20, 30, 40, 50 or 100 contiguous amino acids from the amino acid sequence set forth in figure 1A (SEQ ID NO:1 to 8).

Accordingly, in one aspect of the invention there is provided a fusion partner protein comprising a modified choline binding domain and a heterologous promiscuous T helper epitope, wherein the choline binding domain is selected from the group comprising:

a) the C-terminal domain of LytA as set forth in SEQ ID NO:7;

b) the sequence of SEQ ID NO:8;

c) a peptide sequence comprising an amino acid sequence having at least 85% identity, preferably at least 90% identity, more preferably at least 95% identity, most preferably at least 97-99% identity, to any of SEQ ID NO:1 to 6;

d) a peptide sequence comprising an amino acid sequence having at least 15, 20, 30, 40, 50 or 100 contiguous amino acids from the amino acid sequence of SEQ ID NO:7 or SEQ ID NO:8.

In a most preferred embodiment, the C-LytA extends from amino acid 177-298 which contains a portion of the first repeat and the complete five others, and is set forth in figure 1A.

The second component of the fusion partner, the heterologous T-cell epitope is preferably selected from the group of epitopes that will bind to a number of individuals expressing more than one MHC II molecules in humans. For example, epitopes that are specifically contemplated are P2 and P30 epitopes from tetanus toxoid, Panina – Bordignon

Eur. J. Immunol 19 (12), 2237 (1989). In a preferred embodiment the heterologous T-cell epitope is P2 or P30 from Tetanus toxin.

The P2 epitope has the sequence QYIKANSKFIGITE and corresponds to amino acids 830-843 of the Tetanus toxin.

The P30 epitope (residues 947-967 of Tetanus Toxin) has the sequence FNNFTVSFWLRVPKVSASHLE. The FNNFTV sequence may optionally be deleted. Other universal T epitopes can be derived from the circumsporozoite protein from Plasmodium falciparum – in particular the region 378-398 having the sequence DIEKKIAKMEKASSVFNVVNS (Alexander J, (1994) Immunity 1 (9), p 751-761). Another epitope is derived from Measles virus fusion protein at residue 288-302 having the sequence LSEIKGVIVHRLEGV (Partidos CD, 1990, J. Gen. Virol 71(9) 2099-2105). Yet another epitope is derived from hepatitis B virus surface antigen, in particular amino acids, having the sequence FFLLTRILTIPQSLD.

Another set of epitopes is derived from diphteria toxin. Four of these peptides (amino acids 271-290, 321-340, 331-350, 351-370) map within the T domain of fragment B of the toxin, and the remaining 2 map in the R domain (411-430, 431-450):

PVFAGANYAAWAVNVAQVI

VHHNTEEIVAQSIALSSLMV

QSIALSSLMVAQAIPLVGEL

VDIGFAAYNFVESII NLFQV

QGESGHDIKITAENTPLPIA

GVLLPTIPGKLDVNKSKTHI

(Raju R., Navaneetham D., Okita D., Diethelm-Okita B., McCormick D., Conti-Fine B. M. (1995) Eur. J. Immunol. 25: 3207-14.)

The heterologous T-epitope is preferably fused to C-LytA containing at least 4 repeats, preferably repeat 2 –5 inclusive. One or more subsequent repeats may optionally be fused to the C-terminus of the T-epitope.

Alternatively, the heterologous T-epitope is preferably inserted between two consecutive repeats of C-LytA containing a total of at least 4 repeats, or inserted into one of the repeats of C-LytA containing a total of at least 4 repeats. More preferably, the C-LytA contains 6 repeats and the heterologous epitope is inserted within and at the beginning of the sixth repeat of C-LytA.

The present invention further provides, in other aspects, fusion proteins that comprise at least one polypeptide as described above, as well as polynucleotides encoding

such fusion proteins, typically in the form of pharmaceutical compositions, e.g., vaccine-compositions, comprising a physiologically acceptable carrier and/or an immunostimulant.

Thus a self-protein or other poorly immunogenic protein may be fused to either the N or C terminal end of the resulting fusion partner. Alternatively the self protein or poorly immunogenic protein may be inserted into the fusion partner. In an optional embodiment a histidine tag or at least four, preferably more than 6 histidine residues, may be fused to the alternative end of the poorly immunogenic protein. This would allow for the protein to be purified by affinity chromatography steps, as a histidine tail, typically comprising at least four, preferably six or more residues binds to metal ions and therefore is suitable for metal immobilised metal ion affinity chromatography (IMAC).

Typical constructs would therefore comprise:

- Poorly- immunogenic protein – C-LytA repeats$_{1-4}$ -P_2 epitope (inserted in or replacing C-LytA repeat$_5$)-C-LytA repeat$_6$
- C-LytA repeats$_{1-4}$ -P_2 epitope (inserted in or replacing C-LytA repeat$_5$) – C-LytA repeat$_6$– Poorly immunogenic protein
- Poorly immunogenic protein – C-LytA repeat$_{2-5}$ -P_2 epitope (inserted into C-LytA repeat$_6$)
- C-LytA$_{2-5}$ -P_2 epitope (inserted into C-LytA repeat$_6$)– Poorly immunogenic protein.
- Poorly immunogenic protein C-LytA repeats$_{1-5}$-P_2 epitope- inserted in C-LytA repeat$_6$
- C-LytA repeats$_{1-5}$-P_2 epitope- inserted in C-LytA repeat$_6$- Poorly immunogenic protein
- Poorly immunogenic protein- P_2 epitope inserted into C-LytA repeat$_1$-C-LytA repeats$_{2-5}$
- P_2 epitope inserted into C-LytA repeat$_1$-C-LytA repeats$_{2-5}$- Poorly immunogenic protein
- Poorly immunogenic protein- P_2 epitope inserted into C-LytA repeat$_1$-C-LytA repeats$_{2-6}$
- P_2 epitope inserted into C-LytA repeat$_1$-C-LytA repeats$_{2-6}$- Poorly immunogenic protein
- Poorly immunogenic protein-C-LytA repeat$_1$-P_2 epitope inserted into C-LytA repeat$_2$-C-LytA repeats$_{3-6}$

- C-LytA repeat$_1$-P_2 epitope inserted into C-LytA repeat$_2$-C-LytA repeats$_{3-6}$-Poorly immunogenic protein;

where "inserted into" means at any place into the repeat for example between residue 1 and 2, or between 2 and 3, etc.

The promiscuous T helper epitope may be inserted within a repeat region for example C-LytA repeats $_{2-5}$ _ - C-LytA repeat 6a-P_2 epitope - C-LytA repeat 6b, where the P2 epitope is inserted within the sixth repeat (see figure 7).

In other preferred embodiments the C-terminal end of CPL1 (C-CPL1) may be used as an alternative to C-LytA.

Alternatively, the P2 epitope in the above constructs may be replaced by other promiscuous T epitopes, for example P30. In an embodiment of the invention, two or more promiscuous epitopes are part of the fusion construct. It is however preferred to keep the fusion partner as small as possible, thus limiting the number of potentially interfering CD8+ and B epitopes. Thus the fusion partner is preferably no bigger than 100-140 amino acids, preferably bo bigger than 120 amino acids, typically about 100 amino acid.

The fusion partner of the present invention are preferably fused to a self antigen such as a tumour associated or tissue specific antigens such as those for prostrate, breast, colorectal, lung, pancreatic, ovarian, renal or melanoma cancers. Fragments of said self or tumour antigens are expressly contemplated to be fused to the fusion partner of the invention. Typically the fragment will contain at least 20, preferably 50, more preferably 100 contiguous amino acids of the full-length sequence. Typically such fragments will be devoid of one or more transmembrane domains or may have N-terminal or C-terminal deletions of about 3, 5 , 8, 10, 15, 20, 28 , 33, 50, 54 amino acids. Such fragments will, when suitably presented, be able to generate immune responses that recognise the full length protein.

Particularly illustrative polypeptides of the present invention comprise a sequence of at least 10 contiguous amino acids, preferably 20, more preferably 30, 40, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150, 160, 170, 180 amino acids of a tumour associated or tissue specific protein fused to the fusion partner.

The polypeptides of the invention are immunogenic, i.e., they react detectably within an immunoassay (such as an ELISA or T-cell stimulation assay) with antisera and/or T-cells from a patient with cripto expressing cancer. Screening for immunogenic activity can be performed using techniques well known to the skilled artisan. For example, such

screens can be performed using methods such as those described in Harlow and Lane, *Antibodies: A Laboratory Manual*, Cold Spring Harbor Laboratory, 1988. In one illustrative example, a polypeptide may be immobilised on a solid support and contacted with patient sera to allow binding of antibodies within the sera to the immobilised polypeptide. Unbound sera may then be removed and bound antibodies detected using, for example, ^{125}I-labeled Protein A.

As would be recognised by the skilled artisan, immunogenic portions of tumour associated or tumour specific antigen are also encompassed by the present invention. An "immunogenic portion" as used herein, is a fragment that itself is immunologically reactive (i.e., specifically binds) with the B-cells and/or T-cell surface antigen receptors that recognize the polypeptide. Immunogenic portions may generally be identified using well known techniques, such as those summarized in Paul, *Fundamental Immunology*, 3rd ed., 243-247 (Raven Press, 1993) and references cited therein. Such techniques include screening polypeptides for the ability to react with antigen-specific antibodies, antisera and/or T-cell lines or clones. As used herein, antisera and antibodies are "antigen-specific" if they specifically bind to an antigen (i.e., they react with the protein in an ELISA or other immunoassay, and do not react detectably with unrelated proteins). Such antisera and antibodies may be prepared as described herein, and using well-known techniques.

In one preferred embodiment, an immunogenic portion of a polypeptide is a portion that reacts with antisera and/or T-cells at a level that is not substantially less than the reactivity of the full-length polypeptide (e.g., in an ELISA and/or T-cell reactivity assay). Preferably, the level of immunogenic activity of the immunogenic portion is at least about 50%, preferably at least about 70% and most preferably greater than about 90% of the immunogenicity for the full-length polypeptide. In some instances, preferred immunogenic portions will be identified that have a level of immunogenic activity greater than that of the corresponding full-length polypeptide, e.g., having greater than about 100% or 150% or more immunogenic activity.

In certain other embodiments, illustrative immunogenic portions may include peptides in which an N-terminal leader sequence and/or transmembrane domain have been deleted. Other illustrative immunogenic portions will contain a small N- and/or C-terminal deletion (e.g., about 1-50 amino acids, preferably about 1-30 amino acids, more preferably about 5-15 amino acids), relative to the mature protein.

Exemplary antigens or fragments derived therefrom include MAGE-1, Mage 3 and MAGE 4 or other MAGE antigens such as disclosed in WO99/40188, PRAME, BAGE, LAGE (also known as NY-ESO-1) SAGE and HAGE (WO 99/53061) or GAGE (Robbins and Kawakami, 1996, Current Opinions in Immunology 8, pps 628-636; Van den Eynde et al., International Journal of Clinical & Laboratory Research (submitted 1997); Correale et al. (1997), Journal of the National Cancer Institute 89, p293. Indeed these antigens are expressed in a wide range of tumour types such as melanoma, lung carcinoma, sarcoma and bladder carcinoma.

In a preferred embodiment prostate antigens are utilised, such as Prostate specific antigen (PSA), PAP, PSCA (PNAS 95(4) 1735 –1740 1998), PSMA or antigen known as prostase.

In a particularly preferred embodiment, the prostate antigen is P501S or a fragment thereof. P501S, also named prostein (Xu et al., Cancer Res. 61, 2001, 1563-1568), is known as sequence ID no 113 of WO98/37814 and is a 553 amino acid protein. Immunogenic fragments and portions thereof comprising at least 20, preferably 50, more preferably 100 contiguous amino acids as disclosed in the above referenced patent application and are specifically contemplate by the present invention. Preferred fragments are disclosed in WO 98/50567 (PS108 antigen). Other preferred fragments are amino acids 51-553, 34-553 or 55-553 of the full-length P501S protein.

In particular, construct 1, 2 and 3 (see figure 7) are expressly contemplated, and can be expressed in yeast systems, for example DNA sequences encoding such polypeptides can be expressed in yeast system.

Prostase is a prostate-specific serine protease (trypsin-like), 254 amino acid-long, with a conserved serine protease catalytic triad H-D-S and a amino-terminal pre-propeptide sequence, indicating a potential secretory function (P. Nelson, Lu Gan, C. Ferguson, P. Moss, R. linas, L. Hood & K. Wand, "Molecular cloning and characterisation of prostase, an androgen-regulated serine protease with prostate restricted expression, *In* Proc. Natl. Acad. Sci. USA (1999) 96, 3114-3119). A putative glycosylation site has been described. The predicted structure is very similar to other known serine proteases, showing that the mature polypeptide folds into a single domain. The mature protein is 224 amino acids-long, with one A2 epitope shown to be naturally processed.

Prostase nucleotide sequence and deduced polypeptide sequence and homologous are disclosed in Ferguson, et al. (Proc. Natl. Acad. Sci. USA 1999, 96, 3114-3119) and in International Patent Applications No. WO 98/12302 (and also the corresponding granted

patent US 5,955,306), WO 98/20117 (and also the corresponding granted patents US 5,840,871 and US 5,786,148) (prostate-specific kallikrein) and WO 00/04149 (P703P).

Other prostate specific antigens are known from WO98/37418, and WO/004149. Another is STEAP PNAS 96 14523 14528 7 –12 1999.

Other tumour associated antigens useful in the context of the present invention include: Plu –1 J Biol. Chem 274 (22) 15633 –15645, 1999, HASH –1, HASH-2 (Alders,M. et al., Hum. Mol. Genet. 1997, 6, 859-867), Cripto (Salomon et al Bioessays 199, 21 61 –70,US patent 5654140), Criptin (US patent 5 981 215). Additionally, antigens particularly relevant for vaccines in the therapy of cancer also comprise tyrosinase, telomerase and survivin.

The present invention is also useful in combination with breast cancer antigens such as Her 2/ neu, mammaglobin (US patent 5668267) or those disclosed in WO/00 52165, WO99/33869, WO99/19479, WO 98/45328. Her 2/ neu antigens are disclosed inter alia, in US patent 5,801,005. Preferably the Her 2/ neu comprises the entire extracellular domain (comprising approximately amino acid 1 –645) or fragments thereof and at least an immunogenic portion of or the entire intracellular domain approximately the C terminal 580 amino acids. In particular, the intracellular portion should comprise the phosphorylation domain or fragments thereof. Such constructs are disclosed in WO00/44899. A particularly preferred construct is known as ECD PD a second is known as ECD deltaPD (see WO/00/44899).

The Her 2/ neu as used herein can be derived from rat, mouse or human.

Certain tumour antigens are small peptide antigens (ie less than about 50 amino acids). These antigens can be chemically conjugated to the modified choline binding protein of the present invention.

Exemplary peptides included Mucin derived peptides such as Muc1 see for example US 5744,144 US 5827, 666 WO 8805054, US 4,963,484. Specifically contemplated are Muc 1 derived peptides that comprise at least one repeat unit of the Muc 1 peptide, preferably at least two such repeats and which is recognised by the SM3 antibody (US 6 054 438). Other mucin derived peptides include peptide from Muc 5.

Or said antigen maybe a self peptide hormone such as whole length Gonadotrophin hormone releasing hormone (GnRH, WO 95/20600), a short 10 amino acid long peptide, useful in the treatment of many cancers, or in immunocastration.

Other tumour-specific antigens are suitable to be coupled with the modified Choline binding protein of the present invention include, but are not restricted to tumour-specific gangliosides such as GM2, and GM3.

The covalent coupling of the peptide to modified choline binding protein can be carried out in a manner well known in the art. Thus, for example, for direct covalent coupling it is possible to utilise a carbodiimide, glutaraldehyde or (N-[γ-maleimidobutyryloxy] succinimide ester, utilising common commercially available heterobifunctional linkers such as CDAP and SPDP (using manufacturers instructions). After the coupling reaction, the immunogen can easily be isolated and purified by means of a dialysis method, a gel filtration method, a fractionation method etc.

The present invention also provides a polynucleotide encoding the fusion partner according to the present invention. The invention further relates a polynucleotide that hybridise to the polynucleotide sequence provided herein in figure 1B. In this regard, the invention especially relates to polynucleotides that hybridise under stringent conditions to the polynucleotide described herein. As herein used, the terms "stringent conditions" and "stringent hybridisation conditions" mean hybridisation occurring only if there is at least 95% and preferably at least 97% identity between the sequences. A specific example of stringent hybridization conditions is overnight incubation at 42°C in a solution comprising: 50% formamide, 5x SSC (150mM NaCl, 15mM trisodium citrate), 50 mM sodium phosphate (pH7.6), 5x Denhardt's solution, 10% dextran sulfate, and 20 micrograms/ml of denatured, sheared salmon sperm DNA, followed by washing the hybridisation support in 0.1x SSC at about 65°C. Hybridisation and wash conditions are well known and exemplified in Sambrook, *et al.*, Molecular Cloning: A Laboratory Manual, Second Edition, Cold Spring Harbor, N.Y., (1989), particularly Chapter 11 therein. Solution hybridisation may also be used with the polynucleotide sequences provided by the invention.

The present invention also provides a polynucleotide encoding the polypeptide comprising the fusion partner according to the present invention fused to a tumour associated antigen or fragment thereof.

Such polynucleotide sequences can be inserted into a suitable expression vector and expressed in a suitable host. Vectors may be provided which encode the modified choline binding protein of the invention and which contain a suitable restriction site into which a DNA encoding a poorly immunogenic protein can be inserted to produce a fusion protein.

In other embodiments of the invention, polynucleotide sequences or fragments thereof which encode polypeptide fusions of the invention, may be used in recombinant DNA molecules to direct expression of a polypeptide in appropriate host cells. Due to the inherent degeneracy of the genetic code, other DNA sequences that encode substantially the same or a functionally equivalent amino acid sequence may be produced and these sequences may be used to clone and express a given polypeptide.

As will be understood by those of skill in the art, it may be advantageous in some instances to produce polypeptide-encoding nucleotide sequences possessing non-naturally occurring codons. For example, codons preferred by a particular prokaryotic (for example *E. coli* or yeast) or eukaryotic host can be selected to increase the rate of protein expression, to produce a recombinant RNA transcript having desirable properties, such as a half-life which is longer than that of a transcript generated from the naturally occurring sequence, or to optimise the immune response in humans.

A DNA sequence encoding the fusion proteins or modified choline binding protein of the present invention can be synthesised using standard DNA synthesis techniques, such as by enzymatic ligation as described by D.M. Roberts *et al.* in Biochemistry 1985, *24*, 5090-5098, by chemical synthesis, by *in vitro* enzymatic polymerisation, or by PCR technology utilising for example a heat stable polymerase, or by a combination of these techniques.

Enzymatic polymerisation of DNA may be carried out *in vitro* using a DNA polymerase such as DNA polymerase I (Klenow fragment) or Taq polymerase in an appropriate buffer containing the nucleoside triphosphates dATP, dCTP, dGTP and dTTP as required at a temperature of 10°-37°C, generally in a volume of 50μl or less. Enzymatic ligation of DNA fragments may be carried out using a DNA ligase such as T4 DNA ligase in an appropriate buffer, such as 0.05M Tris (pH 7.4), 0.01M $MgCl_2$, 0.01M dithiothreitol, 1mM spermidine, 1mM ATP and 0.1mg/ml bovine serum albumin, at a temperature of 4°C to ambient, generally in a volume of 50 μl or less. The chemical synthesis of the DNA polymer or fragments may be carried out by conventional phosphotriester, phosphate or phosphoramidite chemistry, using solid phase techniques such as those described in 'Chemical and Enzymatic Synthesis of Gene Fragments - A Laboratory Manual' (ed. H.G. Gassen and A. Lang), Verlag Chemie, Weinheim (1982), or in other scientific publications, for example M.J. Gait, H.W.D. Matthes, M. Singh, B.S. Sproat, and R.C. Titmas, Nucleic Acids Research, 1982, 10, 6243; B.S. Sproat, and W. Bannwarth, Tetrahedron Letters,

1983, 24, 5771; M.D. Matteucci and M.H. Caruthers, Tetrahedron Letters, 1980, 21, 719; M.D. Matteucci and M.H. Caruthers, Journal of the American Chemical Society, 1981, 103, 3185; S.P. Adams *et al.*, Journal of the American Chemical Society, 1983, 105, 661; N.D. Sinha, J. Biernat, J. McMannus, and H. Koester, Nucleic Acids Research, 1984, 12, 4539; and H.W.D. Matthes *et al.*, EMBO Journal, 1984, 3, 801.

The process of the invention may be performed by conventional recombinant techniques such as described in Maniatis *et al.*, Molecular Cloning - A Laboratory Manual; Cold Spring Harbor, 1982-1989.

In particular, the process may comprise the steps of :

i) preparing a replicable or integrating expression vector capable, in a host cell, of expressing a DNA polymer comprising a nucleotide sequence that encodes the protein or an immunogenic derivative thereof

ii) transforming a host cell with said vector

iii) culturing said transformed host cell under conditions permitting expression of said DNA polymer to produce said protein; and

iv) recovering said protein

The term 'transforming' is used herein to mean the introduction of foreign DNA into a host cell. This can be achieved for example by transformation, transfection or infection with an appropriate plasmid or viral vector using e.g. conventional techniques as described in Genetic Engineering; Eds. S.M. Kingsman and A.J. Kingsman; Blackwell Scientific Publications; Oxford, England, 1988. The term 'transformed' or 'transformant' will hereafter apply to the resulting host cell containing and expressing the foreign gene of interest.

The expression vectors are novel and also form part of the invention.

The replicable expression vectors may be prepared in accordance with the invention, by cleaving a vector compatible with the host cell to provide a linear DNA segment having an intact replicon, and combining said linear segment with one or more DNA molecules which, together with said linear segment encode the desired product, such as the DNA polymer encoding the protein of the invention, or derivative thereof, under ligating conditions.

Thus, the DNA polymer may be performed or formed during the construction of the vector, as desired.

The choice of vector will be determined in part by the host cell, which may be prokaryotic or eukaryotic but are preferably *E. coli*, yeast or CHO cells. Suitable vectors include plasmids, bacteriophages, cosmids and recombinant viruses. Expression and cloning vectors preferably contain a selectable marker such that only the host cells expressing the marker will survive under selective conditions. Selection genes include but are not limited to the one encoding protein that confer a resistance to ampicillin, tetracyclin or kanamycin. Expression vectors also contain control sequences which are compatible with the designated host. For example, expression control sequences for *E. coli*, and more generally for prokaryotes, include promoters and ribosome binding sites. Promoter sequences may be naturally occurring, such as the β-lactamase (penicillinase) (Weissman 1981, *In Interferon 3* (ed. L. Gresser), lactose (lac) (Chang et al. Nature, 1977, 198: 1056) and tryptophan (trp) (Goeddel et al. Nucl. Acids Res. 1980, 8, 4057) and lambda-derived P_L promoter system. In addition, synthetic promoters which do not occur in nature also function as bacterial promoters. This is the case for example for the tac synthetic hybrid promoter which is derived from sequences of the trp and lac promoters (De Boer et al., Proc. Natl Acad Sci. USA 1983, 80, 21-26). These systems are particularly suitable with *E. coli*.

Yeast compatible vectors also carry markers that allow the selection of successful transformants by conferring prototrophy to auxotrophic mutants or resistance to heavy metals on wild-type strains. Expression control sequences for yeast vectors include promoters for glycolytic enzymes (Hess et al., J. Adv. Enzyme Reg. 1968, 7, 149), PHO5 gene encoding acid phosphatase, CUP1 gene, ARG3 gene, GAL genes promoters and synthetic promoter sequences.. Other control elements useful in yeast expression are terminators and mRNA leader sequences. The 5' coding sequence is particularly useful since it typically encodes a signal peptide comprised of hydrophobic amino acids which direct the secretion of the protein from the cell. Suitable signal sequences can be encoded by genes for secreted yeast proteins such as the yeast invertase gene and the a-factor gene, acid phosphatase, killer toxin, the alpha-mating factor gene and recently the heterologous inulinase signal sequence derived from INU1A gene of Kluyveromyces marxianus.. Suitable vectors have been developed for expression in *Pichia pastoris* and *Saccharomyces cerevisiae*.

Ã variety of *P. pastoris* expression vectors are available based on various inducible or constitutive promoters (Cereghino and Cregg, FEMS Microbiol. Rev. 2000,24:45-66). For the production of cytosolic and secreted proteins,the most commonly used P. pastoris vectors contain the very strong and tightly regulated alcohol oxidase (AOX1) promoter. The vectors also contain the *P. pastoris* histidinol dehydrogenase (HIS4) gene for selection in his4 hosts. Secretion of foreign protein require the presence of a signal sequence and the *S. cerevisiae* prepro alpha mating factor signal sequence has been widly and successfully used in Pichia expression system. Expression vectors are integrated into the P. pastoris genome to maximize the stability of expression strains. As in *S.cerevisiae*, cleavage of a *P.pastoris* expression vector within a sequence shared by the host genome (AOX1 or HIS4) stimulates homologous recombination events that efficiently target integration of the vector to that genomic locus. In general, a recombinant strain that contains multiple integrated copies of an expression cassette can yield more heterologous protein than single-copy strain. The most effective way to obtain high copy number transformants requires the transformation of Pichia recipient strain by the sphaeroplast technique (Cregg et all 1985, Mol.Cell.Biol. 5: 3376-3385) .

The preparation of the replicable expression vector may be carried out conventionally with appropriate enzymes for restriction, polymerisation and ligation of the DNA, by procedures described in, for example, Maniatis *et al.* cited above.

The recombinant host cell is prepared, in accordance with the invention, by transforming a host cell with a replicable expression vector of the invention under transforming conditions. Suitable transforming conditions are conventional and are described in, for example, Maniatis *et al.* cited above, or "DNA Cloning" Vol. II, D.M. Glover ed., IRL Press Ltd, 1985.

The choice of transforming conditions depends upon the choice of the host cell to be transformed. For example, in vivo transformation using a live viral vector as the transforming agent for the polynucleotides of the invention is described above. Bacterial transformation of a host such as *E. coli* may be done by direct uptake of the polynucleotides (which may be expression vectors containing the desired sequence) after the host has been treated with a solution of $CaCl_2$ (Cohen *et al.*, Proc. Nat. Acad. Sci., 1973, *69*, 2110) or with a solution comprising a mixture of rubidium chloride (RbCl), $MnCl_2$, potassium acetate and glycerol, and then with 3-[N-morpholino]-propane-sulphonic acid, RbCl and glycerol or by electroporation. Transformation of lower eukaryotic organisms such as yeast cells in culture by direct uptake may be carried out for example by using the method of

Hinnen et al (Proc. Natl. Acad. Sci. 1978, 75 : 1929-1933). Mammalian cells in culture may be transformed using the calcium phosphate co-precipitation of the vector DNA onto the cells (Graham & Van der Eb, Virology 1978, 52, 546). Other methods for introduction of polynucleotides into mammalian cells include dextran mediated transfection, polybrene mediated transfection, protoplast fusion, electroporation, encapsulation of the polynucleotide(s) into liposomes, and direct micro-injection of the polynucleotides into nuclei.

The invention also extends to a host cell transformed with a nucleic acid encoding the protein of the invention or a replicable expression vector of the invention.

Culturing the transformed host cell under conditions permitting expression of the DNA polymer is carried out conventionally, as described in, for example, Maniatis *et al.* and "DNA Cloning" cited above. Thus, preferably the cell is supplied with nutrient and cultured at a temperature below 50°C, preferably between 25°C and 42°C, more preferably between 25°C and 35°C, most preferably at 30°C. The incubation time may vary from a few minutes to a few hours, according to the proportion of the polypeptide in the bacterial cell, as assessed by SDS-PAGE or Western blot.

The product may be recovered by conventional methods according to the host cell and according to the localisation of the expression product (intracellular or secreted into the culture medium or into the cell periplasm). Thus, where the host cell is bacterial, such as *E. coli* it may, for example, be lysed physically, chemically or enzymatically and the protein product isolated from the resulting lysate. Where the host cell is mammalian, the product may generally be isolated from the nutrient medium or from cell free extracts. Where the host cell is a yeast such as *Saccharomyces cerevisiae* or *Pichia pastoris*, the product may generally be isolated from from lysed cells or from the culture medium, and then further purified using conventional techniques. The specificity of the expression system may be assessed by western blot or by ELISA using an antibody directed against the polypeptide of interest.

Conventional protein isolation techniques include selective precipitation, adsorption chromatography, and affinity chromatography including a monoclonal antibody affinity column. When the proteins of the present invention are expressed with a histidine tail (His tag), they can easily be purified by affinity chromatography using an ion metal affinity chromatography column (IMAC) column.The metal ion, may be any suitable ion for example zinc, nickel, iron, magnesium or copper, but is preferably zinc or nickel.

Preferably the IMAC buffer contains detergent, preferably an anionic detergent such as

SDS, more preferably a non-ionic detergent such as Tween 80, or a zwitterionic detergent such as Empigen BB, as this may result in lower levels of endotoxin in the final product.

Further chromatographic steps include for example a Q-Sepharose step that may be operated either before of after the IMAC column. Preferably the pH is in the range of 7.5 to 10, more preferably from 7.5 to 9.5, optimally between 8 and 9.

The proteins of the invention can thus be purified according to the following protocol. After cell disruption, cell extracts containing the protein can be solubilised in a pH 8.5 Tris buffer containing urea (8.0 M for example), and SDS (from 0.5% to 1% for example). After centrifugation, the resulting supernatant may then be loaded onto on to an IMAC (Nickel) Sepharose FF column equilibrated with a pH 8.5 Tris buffer. The column may then be washed with a high salt containing buffer (eg 0.75 – 1.5m NaCl, 15 mM pH 8.5 Tris buffer). The column may optionally then be washed again with phosphate buffer without salt. The proteins of the invention may be eluated from the column with an imidazole-containing buffered solution. The proteins can then be submitted to an additional chromatographic step, such as to an anion exchange chromatography (Q Sepharose for example). .

The proteins of the present invention are provided either soluble in a liquid form or in a lyophilised form, which is the preferred form. It is generally expected that each human dose will comprise 1 to 1000 µg of protein, and preferably 30-300 µg. The purification process can also include a carboxyamidation step whereby the protein is first reduced in the presence of Glutathion and then carboxymethylated in the presence of iodoacetamide. This step offers the advantage of controling the oxidative aggregation of the molecule with itself or with host cell protein contaminants through covalent bridging with disulphide bonds.

The present invention also provides pharmaceutical and immunogenic compositions comprising a protein of the present invention in a pharmaceutically acceptable excipient. A preferred vaccine composition comprises at least a protein according to the invention. Said protein has, preferably, blocked thiol groups and is highly purified, e.g. has less than 5% host cell contamination. Such vaccine may optionally contain one or more other tumour-associated antigen and derivatives. For example, suitable other associated antigen include prostase, PAP-1, PSA (prostate specific antigen), PSMA (prostate-specific membrane antigen), PSCA (Prostate Stem Cell Antigen), STEAP.

In another embodiment, illustrative immunogenic compositions, such as for example vaccine compositions, of the present invention comprise DNA encoding one or more of the fusion polypeptides as described above, such that the fusion polypeptide is

generated *in situ*. As noted above, the polynucleotide may be administered within any of a variety of delivery systems known to those of ordinary skill in the art. Indeed, numerous gene delivery techniques are well known in the art, such as those described by Rolland, *Crit. Rev. Therap. Drug Carrier Systems 15*:143-198, 1998, and references cited therein. Appropriate polynucleotide expression systems will, of course, contain the necessary regulatory DNA regulatory sequences for expression in a patient (such as a suitable promoter and terminating signal). Alternatively, bacterial delivery systems may involve the administration of a bacterium (such as *Bacillus-Calmette-Guerrin*) that expresses an immunogenic portion of the polypeptide on its cell surface or secretes such an epitope.

Therefore, in certain embodiments, polynucleotides encoding immunogenic polypeptides described herein are introduced into suitable mammalian host cells for expression using any of a number of known viral-based systems. In one illustrative embodiment, retroviruses provide a convenient and effective platform for gene delivery systems. A selected nucleotide sequence encoding a polypeptide of the present invention can be inserted into a vector and packaged in retroviral particles using techniques known in the art. The recombinant virus can then be isolated and delivered to a subject. A number of illustrative retroviral systems have been described (e.g., U.S. Pat. No. 5,219,740; Miller and Rosman (1989) BioTechniques 7:980-990; Miller, A. D. (1990) Human Gene Therapy 1:5-14; Scarpa et al. (1991) Virology 180:849-852; Burns et al. (1993) Proc. Natl. Acad. Sci. USA 90:8033-8037; and Boris-Lawrie and Temin (1993) Cur. Opin. Genet. Develop. 3:102-109.

In addition, a number of illustrative adenovirus-based systems have also been described. Unlike retroviruses which integrate into the host genome, adenoviruses persist extrachromosomally thus minimizing the risks associated with insertional mutagenesis (Haj-Ahmad and Graham (1986) J. Virol. 57:267-274; Bett et al. (1993) J. Virol. 67:5911-5921; Mittereder et al. (1994) Human Gene Therapy 5:717-729; Seth et al. (1994) J. Virol. 68:933-940; Barr et al. (1994) Gene Therapy 1:51-58; Berkner, K. L. (1988) BioTechniques 6:616-629; and Rich et al. (1993) Human Gene Therapy 4:461-476).

Various adeno-associated virus (AAV) vector systems have also been developed for polynucleotide delivery. AAV vectors can be readily constructed using techniques well known in the art. See, e.g., U.S. Pat. Nos. 5,173,414 and 5,139,941; International Publication Nos. WO 92/01070 and WO 93/03769; Lebkowski et al. (1988) Molec. Cell. Biol. 8:3988-3996; Vincent et al. (1990) Vaccines 90 (Cold Spring Harbor Laboratory

Press); Carter, B. J. (1992) Current Opinion in Biotechnology 3:533-539; Muzyczka, N. (1992) Current Topics in Microbiol. and Immunol. 158:97-129; Kotin, R. M. (1994) Human Gene Therapy 5:793-801; Shelling and Smith (1994) Gene Therapy 1:165-169; and Zhou et al. (1994) J. Exp. Med. 179:1867-1875.

Additional viral vectors useful for delivering the nucleic acid molecules encoding polypeptides of the present invention by gene transfer include those derived from the pox family of viruses, such as vaccinia virus and avian poxvirus. By way of example, vaccinia virus recombinants expressing the novel molecules can be constructed as follows. The DNA encoding a polypeptide is first inserted into an appropriate vector so that it is adjacent to a vaccinia promoter and flanking vaccinia DNA sequences, such as the sequence encoding thymidine kinase (TK). This vector is then used to transfect cells which are simultaneously infected with vaccinia. Homologous recombination serves to insert the vaccinia promoter plus the gene encoding the polypeptide of interest into the viral genome. The resulting TK.sup.(-) recombinant can be selected by culturing the cells in the presence of 5-bromodeoxyuridine and picking viral plaques resistant thereto.

A vaccinia-based infection/transfection system can be conveniently used to provide for inducible, transient expression or coexpression of one or more polypeptides described herein in host cells of an organism. In this particular system, cells are first infected in vitro with a vaccinia virus recombinant that encodes the bacteriophage T7 RNA polymerase. This polymerase displays exquisite specificity in that it only transcribes templates bearing T7 promoters. Following infection, cells are transfected with the polynucleotide or polynucleotides of interest, driven by a T7 promoter. The polymerase expressed in the cytoplasm from the vaccinia virus recombinant transcribes the transfected DNA into RNA which is then translated into polypeptide by the host translational machinery. The method provides for high level, transient, cytoplasmic production of large quantities of RNA and its translation products. See, e.g., Elroy-Stein and Moss, Proc. Natl. Acad. Sci. USA (1990) 87:6743-6747; Fuerst et al. Proc. Natl. Acad. Sci. USA (1986) 83:8122-8126.

Alternatively, avipoxviruses, such as the fowlpox and canarypox viruses, can also be used to deliver the coding sequences of interest. Recombinant avipox viruses, expressing immunogens from mammalian pathogens, are known to confer protective immunity when administered to non-avian species. The use of an Avipox vector is particularly desirable in human and other mammalian species since members of the Avipox genus can only productively replicate in susceptible avian species and therefore are not infective in

mammalian cells. Methods for producing recombinant Avipoxviruses are known in the art and employ genetic recombination, as described above with respect to the production of vaccinia viruses. See, e.g., WO 91/12882; WO 89/03429; and WO 92/03545.

Any of a number of alphavirus vectors can also be used for delivery of polynucleotide compositions of the present invention, such as those vectors described in U.S. Patent Nos. 5,843,723; 6,015,686; 6,008,035 and 6,015,694. Certain vectors based on Venezuelan Equine Encephalitis (VEE) can also be used, illustrative examples of which can be found in U.S. Patent Nos. 5,505,947 and 5,643,576.

In another embodiment of the invention, a polynucleotide is administered/delivered as "naked" DNA, for example as described in Ulmer et al., *Science 259*:1745-1749, 1993 and reviewed by Cohen, *Science 259*:1691-1692, 1993. The uptake of naked DNA may be increased by coating the DNA onto biodegradable beads, which are efficiently transported into the cells.

The fusion proteins of the invention can also be formulated as a phamaceutical composition, e.g. as a vaccine.

The fusion proteins of the present invention are provided preferably at least 80% pure more preferably 90% pure as visualised by SDS PAGE. Preferably the proteins appear as a single band by SDS PAGE.

The present invention also provides pharmaceutical composition comprising a fusion protein of the present invention in a pharmaceutically acceptable excipient. Accordingly there is also provided a process for the preparation of a immunogenic composition according to the present invention, comprising admixing the fusion protein of the invention or a the encoding polynucleotide with a suitable adjuvant, diluent or other pharmaceutically acceptable carrier

Vaccine preparation is generally described in Vaccine Design ("The subunit and adjuvant approach" (eds. Powell M.F. & Newman M.J). (1995) Plenum Press New York). Encapsulation within liposomes is described by Fullerton, US Patent 4,235,877.

The fusion proteins of the present invention are preferably adjuvanted in the vaccine formulation of the invention. Certain adjuvants are commercially available as, for example, Freund's Incomplete Adjuvant and Complete Adjuvant (Difco Laboratories, Detroit, MI); Merck Adjuvant 65 (Merck and Company, Inc., Rahway, NJ); AS-2 (SmithKline Beecham, Philadelphia, PA); aluminum salts such as aluminum hydroxide gel (alum) or aluminum

phosphate; salts of calcium, iron or zinc; an insoluble suspension of acylated tyrosine; acylated sugars; cationically or anionically derivatised polysaccharides; polyphosphazenes; biodegradable microspheres; monophosphoryl lipid A and quil A. Cytokines, such as GM-CSF, interleukin-2, -7, -12, and other like growth factors, may also be used as adjuvants.

Within certain embodiments of the invention, the adjuvant composition is preferably one that induces an immune response predominantly of the Th1 type. High levels of Th1-type cytokines (*e.g.*, IFN-γ, TNFα, IL-2 and IL-12) tend to favor the induction of cell mediated immune responses to an administered antigen. In contrast, high levels of Th2-type cytokines (*e.g.*, IL-4, IL-5, IL-6 and IL-10) tend to favor the induction of humoral immune responses. Following application of a vaccine as provided herein, a patient will support an immune response that includes Th1- and Th2-type responses. Within a preferred embodiment, in which a response is predominantly Th1-type, the level of Th1-type cytokines will increase to a greater extent than the level of Th2-type cytokines. The levels of these cytokines may be readily assessed using standard assays. For a review of the families of cytokines, see Mosmann and Coffman, *Ann. Rev. Immunol. 7*:145-173, 1989.

Preferred TH-1 inducing adjuvants are selected from the group of adjuvants comprising: 3D-MPL, QS21, a mixture of QS21 and cholesterol, and a CpG oligonucleotide or a mixture of two or more said adjuvants. Certain preferred adjuvants for eliciting a predominantly Th1-type response include, for example, a combination of monophosphoryl lipid A, preferably 3-de-O-acylated monophosphoryl lipid A, together with an aluminum salt. MPL® adjuvants are available from Corixa Corporation (Seattle, WA; *see*, for example, US Patent Nos. 4,436,727; 4,877,611; 4,866,034 and 4,912,094). CpG-containing oligonucleotides (in which the CpG dinucleotide is unmethylated) also induce a predominantly Th1 response. Such oligonucleotides are well known and are described, for example, in WO 96/02555, WO 99/33488 and U.S. Patent Nos. 6,008,200 and 5,856,462. Immunostimulatory DNA sequences are also described, for example, by Sato et al., *Science 273*:352, 1996. Another preferred adjuvant comprises a saponin, such as Quil A, or derivatives thereof, including QS21 and QS7 (Aquila Biopharmaceuticals Inc., Framingham, MA); Escin; Digitonin; or *Gypsophila* or *Chenopodium quinoa* saponins . Other preferred formulations include more than one saponin in the adjuvant combinations of the present invention, for example combinations of at least two of the following group comprising QS21, QS7, Quil A, β-escin, or digitonin.

Alternatively the saponin formulations may be combined with vaccine vehicles composed of chitosan or other polycationic polymers, polylactide and polylactide-co-glycolide particles, poly-N-acetyl glucosamine-based polymer matrix, particles composed of polysaccharides or chemically modified polysaccharides, liposomes and lipid-based particles, particles composed of glycerol monoesters, etc. The saponins may also be formulated in the presence of cholesterol to form particulate structures such as liposomes or ISCOMs. Furthermore, the saponins may be formulated together with a polyoxyethylene ether or ester, in either a non-particulate solution or suspension, or in a particulate structure such as a paucilamelar liposome or ISCOM. The saponins may also be formulated with excipients such as Carbopol^{R} to increase viscosity, or may be formulated in a dry powder form with a powder excipient such as lactose.

In one preferred embodiment, the adjuvant system includes the combination of a monophosphoryl lipid A and a saponin derivative, such as the combination of QS21 and 3D-MPL® adjuvant, as described in WO 94/00153, or a less reactogenic composition where the QS21 is quenched with cholesterol, as described in WO 96/33739. Other preferred formulations comprise an oil-in-water emulsion and tocopherol. Another particularly preferred adjuvant formulation employing QS21, 3D-MPL® adjuvant and tocopherol in an oil-in-water emulsion is described in WO 95/17210.

Another enhanced adjuvant system involves the combination of a CpG-containing oligonucleotide and a saponin derivative particularly the combination of CpG and QS21 as disclosed in WO 00/09159. Preferably the formulation additionally comprises an oil in water emulsion and tocopherol.

In a yet further embodiment the present invention provides an immunogenic composition comprising a fusion protein according to the invention, and further comprising D3-MPL, a saponin preferably QS21 and a CpG oligonucleotide, optionally formulated in an oil in water emulsion.

Additional illustrative adjuvants for use in the pharmaceutical compositions of the invention include Montanide ISA 720 (Seppic, France), SAF (Chiron, California, United States), ISCOMS (CSL), MF-59 (Chiron), the SBAS series of adjuvants (*e.g.*, SBAS-2 or SBAS-4, available from SmithKline Beecham, Rixensart, Belgium), **Detox (Enhanzyn®)** (Corixa, Hamilton, MT), RC-529 (Corixa, Hamilton, MT) and other aminoalkyl glucosaminide 4-phosphates (AGPs), such as those described in pending U.S. Patent Application Serial Nos. 08/853,826 and 09/074,720, the disclosures of which are

incorporated herein by reference in their entireties, and polyoxyethylene ether adjuvants such as those described in WO 99/52549A1.

Other preferred adjuvants include adjuvant molecules of the general formula (I):

$HO(CH_2CH_2O)_n$-A-R

Wherein, n is 1-50, A is a bond or –C(O)-, R is C_{1-50} alkyl or Phenyl C_{1-50} alkyl.

One embodiment of the present invention consists of a vaccine formulation comprising a polyoxyethylene ether of general formula (I), wherein *n* is between 1 and 50, preferably 4-24, most preferably 9; the *R* component is C_{1-50}, preferably C_4-C_{20} alkyl and most preferably C_{12} alkyl, and *A* is a bond. The concentration of the polyoxyethylene ethers should be in the range 0.1-20%, preferably from 0.1-10%, and most preferably in the range 0.1-1%. Preferred polyoxyethylene ethers are selected from the following group: polyoxyethylene-9-lauryl ether, polyoxyethylene-9-steoryl ether, polyoxyethylene-8-steoryl ether, polyoxyethylene-4-lauryl ether, polyoxyethylene-35-lauryl ether, and polyoxyethylene-23-lauryl ether. Polyoxyethylene ethers such as polyoxyethylene lauryl ether are described in the Merck index (12^{th} edition: entry 7717). These adjuvant molecules are described in WO 99/52549.

The polyoxyethylene ether according to the general formula (I) above may, if desired, be combined with another adjuvant. For example, a preferred adjuvant combination is preferably with CpG as described in the pending UK patent application GB 9820956.2.

Within further aspects, the present invention provides methods for stimulating an immune response in a patient, preferably a T cell response in a human patient, comprising administering a pharmaceutical composition described herein. The patient may be afflicted with lung or colon cancer or colorectal cancer or breast cancer, in which case the methods provide treatment for the disease, or patient considered at risk for such a disease may be treated prophylactically.

Within further aspects, the present invention provides methods for inhibiting the development of a cancer in a patient, comprising administering to a patient a pharmaceutical composition as recited above. The patient may be afflicted with, for example, sarcoma, prostate, ovarian, bladder, lung, colon, colorectal or breast cancer, in which case the methods provide treatment for the disease, or patient considered at risk for such a disease may be treated prophylactically.

The present invention further provides, within other aspects, methods for removing tumor cells from a biological sample, comprising contacting a biological sample with T

cells that specifically react with a polypeptide of the present invention, wherein the step of contacting is performed under conditions and for a time sufficient to permit the removal of cells expressing the protein from the sample.

Within related aspects, methods are provided for inhibiting the development of a cancer in a patient, comprising administering to a patient a biological sample treated as described above.

Methods are further provided, within other aspects, for stimulating and/or expanding T cells specific for a polypeptide of the present invention, comprising contacting T cells with one or more of: (i) a polypeptide as described above; (ii) a polynucleotide encoding such a polypeptide; and/or (iii) an antigen presenting cell that expresses such a polypeptide; under conditions and for a time sufficient to permit the stimulation and/or expansion of T cells. Isolated T cell populations comprising T cells prepared as described above are also provided.

Within further aspects, the present invention provides methods for inhibiting the development of a cancer in a patient, comprising administering to a patient an effective amount of a T cell population as described above.

The present invention further provides methods for inhibiting the development of a cancer in a patient, comprising the steps of: (a) incubating CD4+ and/or CD8+ T cells isolated from a patient with one or more of: (i) a polypeptide disclosed herein; (ii) a polynucleotide encoding such a polypeptide; and (iii) an antigen-presenting cell that expressed such a polypeptide; and (b) administering to the patient an effective amount of the proliferated T cells, and thereby inhibiting the development of a cancer in the patient. Proliferated cells may, but need not, be cloned prior to administration to the patient.

According to another embodiment of this invention, an immunogenic composition described herein is delivered to a host via antigen presenting cells (APCs), such as dendritic cells, macrophages, B cells, monocytes and other cells that may be engineered to be efficient APCs. Such cells may, but need not, be genetically modified to increase the capacity for presenting the antigen, to improve activation and/or maintenance of the T cell response, to have anti-tumor effects *per se* and/or to be immunologically compatible with the receiver (*i.e.*, matched HLA haplotype). APCs may generally be isolated from any of a variety of biological fluids and organs, including tumor and peritumoral tissues, and may be autologous, allogeneic, syngeneic or xenogeneic cells.

Certain preferred embodiments of the present invention use dendritic cells or progenitors thereof as antigen-presenting cells. Dendritic cells are highly potent APCs (Banchereau and Steinman, *Nature 392*:245-251, 1998) and have been shown to be effective as a physiological adjuvant for eliciting prophylactic or therapeutic antitumor immunity (*see* Timmerman and Levy, *Ann. Rev. Med. 50*:507-529, 1999). In general, dendritic cells may be identified based on their typical shape (stellate *in situ*, with marked cytoplasmic processes (dendrites) visible *in vitro*), their ability to take up, process and present antigens with high efficiency and their ability to activate naïve T cell responses. Dendritic cells may, of course, be engineered to express specific cell-surface receptors or ligands that are not commonly found on dendritic cells *in vivo* or *ex vivo*, and such modified dendritic cells are contemplated by the present invention. As an alternative to dendritic cells, secreted vesicles antigen-loaded dendritic cells (called exosomes) may be used within a vaccine (*see* Zitvogel et al., *Nature Med. 4*:594-600, 1998).

Dendritic cells and progenitors may be obtained from peripheral blood, bone marrow, tumor-infiltrating cells, peritumoral tissues-infiltrating cells, lymph nodes, spleen, skin, umbilical cord blood or any other suitable tissue or fluid. For example, dendritic cells may be differentiated *ex vivo* by adding a combination of cytokines such as GM-CSF, IL-4, IL-13 and/or TNFα to cultures of monocytes harvested from peripheral blood. Alternatively, CD34 positive cells harvested from peripheral blood, umbilical cord blood or bone marrow may be differentiated into dendritic cells by adding to the culture medium combinations of GM-CSF, IL-3, TNFα, CD40 ligand, LPS, flt3 ligand and/or other compound(s) that induce differentiation, maturation and proliferation of dendritic cells.

Dendritic cells are conveniently categorized as "immature" and "mature" cells, which allows a simple way to discriminate between two well characterized phenotypes. However, this nomenclature should not be construed to exclude all possible intermediate stages of differentiation. Immature dendritic cells are characterized as APC with a high capacity for antigen uptake and processing, which correlates with the high expression of Fcγ receptor and mannose receptor. The mature phenotype is typically characterized by a lower expression of these markers, but a high expression of cell surface molecules responsible for T cell activation such as class I and class II MHC, adhesion molecules (*e.g.*, CD54 and CD11) and costimulatory molecules (*e.g.*, CD40, CD80, CD86 and 4-1BB).

APCs may generally be transfected with a polynucleotide of the invention (or portion or other variant thereof) such that the encoded polypeptide, or an immunogenic portion thereof, is expressed on the cell surface. Such transfection may take place *ex vivo*, and a pharmaceutical composition comprising such transfected cells may then be used for therapeutic purposes, as described herein. Alternatively, a gene delivery vehicle that targets a dendritic or other antigen presenting cell may be administered to a patient, resulting in transfection that occurs *in vivo*. *In vivo* and *ex vivo* transfection of dendritic cells, for example, may generally be performed using any methods known in the art, such as those described in WO 97/24447, or the gene gun approach described by Mahvi et al., *Immunology and cell Biology 75*:456-460, 1997. Antigen loading of dendritic cells may be achieved by incubating dendritic cells or progenitor cells with the tumor polypeptide, DNA (naked or within a plasmid vector) or RNA; or with antigen-expressing recombinant bacterium or viruses (*e.g.*, vaccinia, fowlpox, adenovirus or lentivirus vectors). Prior to loading, the polypeptide may be covalently conjugated to an immunological partner that provides T cell help (*e.g.*, a carrier molecule). Alternatively, a dendritic cell may be pulsed with a non-conjugated immunological partner, separately or in the presence of the polypeptide.

Definitions

Also provided by the invention are methods for the analysis of character sequences or strings, particularly genetic sequences or encoded protein sequences. Preferred methods of sequence analysis include, for example, methods of sequence homology analysis, such as identity and similarity analysis, DNA, RNA and protein structure analysis, sequence assembly, cladistic analysis, sequence motif analysis, open reading frame determination, nucleic acid base calling, codon usage analysis, nucleic acid base trimming, and sequencing chromatogram peak analysis.

A computer based method is provided for performing homology identification. This method comprises the steps of: providing a first polynucleotide sequence comprising the sequence of a polynucleotide of the invention in a computer readable medium; and comparing said first polynucleotide sequence to at least one second polynucleotide or polypeptide sequence to identify homology.

A computer based method is also provided for performing homology identification, said method comprising the steps of: providing a first polypeptide sequence comprising the sequence of a polypeptide of the invention in a computer readable medium; and comparing said first polypeptide sequence to at least one second polynucleotide or polypeptide sequence to identify homology.

All publications and references, including but not limited to patents and patent applications, cited in this specification are herein incorporated by reference in their entirety as if each individual publication or reference were specifically and individually indicated to be incorporated by reference herein as being fully set forth. Any patent application to which this application claims priority is also incorporated by reference herein in its entirety in the manner described above for publications and references.

"Identity," as known in the art, is a relationship between two or more polypeptide sequences or two or more polynucleotide sequences, as the case may be, as determined by comparing the sequences. In the art, "identity" also means the degree of sequence relatedness between polypeptide or polynucleotide sequences, as the case may be, as determined by the match between strings of such sequences. "Identity" can be readily calculated by known methods, including but not limited to those described in (*Computational Molecular Biology*, Lesk, A.M., ed., Oxford University Press, New York, 1988; *Biocomputing: Informatics and Genome Projects*, Smith, D.W., ed., Academic Press, New York, 1993; *Computer Analysis of Sequence Data*, Part I, Griffin, A.M., and Griffin, H.G., eds., Humana Press, New Jersey, 1994; *Sequence Analysis in Molecular Biology*, von Heine, G., Academic Press, 1987; and *Sequence Analysis Primer*, Gribskov, M. and Devereux, J., eds., M Stockton Press, New York, 1991; and Carillo, H., and Lipman, D., SIAM *J. Applied Math., 48:* 1073 (1988). Methods to determine identity are designed to give the largest match between the sequences tested. Moreover, methods to determine identity are codified in publicly available computer programs. Computer program methods to determine identity between two sequences include, but are not limited to, the GAP program in the GCG program package (Devereux, J., et al., *Nucleic Acids Research 12(1):* 387 (1984)), BLASTP, BLASTN (Altschul, S.F. et al., *J. Molec. Biol. 215:* 403-410 (1990), and FASTA(Pearson and Lipman Proc. Natl. Acad. Sci. USA 85; 2444-2448 (1988). The BLAST family of programs is publicly available from NCBI and other sources (*BLAST Manual*, Altschul, S., *et al.*, NCBI NLM NIH Bethesda, MD 20894; Altschul, S., *et al., J. Mol. Biol. 215*: 403-410 (1990). The well known Smith Waterman algorithm may also be used to determine identity.

Parameters for polypeptide sequence comparison include the following:
Algorithm: Needleman and Wunsch, J. Mol Biol. 48: 443-453 (1970)
Comparison matrix: BLOSSUM62 from Henikoff and Henikoff,
Proc. Natl. Acad. Sci. USA. 89:10915-10919 (1992)
Gap Penalty: 8
Gap Length Penalty: 2
A program useful with these parameters is publicly available as the "gap" program from Genetics Computer Group, Madison WI. The aforementioned parameters are the default parameters for peptide comparisons (along with no penalty for end gaps).

Parameters for polynucleotide comparison include the following:
Algorithm: Needleman and Wunsch, J. Mol Biol. 48: 443-453 (1970)
Comparison matrix: matches = +10, mismatch = 0
Gap Penalty: 50
Gap Length Penalty: 3
Available as: The "gap" program from Genetics Computer Group, Madison WI. These are the default parameters for nucleic acid comparisons.

A preferred meaning for "identity" for polynucleotides and polypeptides, as the case may be, are provided in (1) and (2) below.

(1) Polynucleotide embodiments further include an isolated polynucleotide comprising a polynucleotide sequence having at least a 50, 60, 70, 80, 85, 90, 95, 97 or 100% identity to any of the reference sequences of SEQ ID NO:9 to SEQ ID NO:16, wherein said polynucleotide sequence may be identical to any the reference sequences of SEQ ID NO:9 to SEQ ID NO:16 or may include up to a certain integer number of nucleotide alterations as compared to the reference sequence, wherein said alterations are selected from the group consisting of at least one nucleotide deletion, substitution, including transition and transversion, or insertion, and wherein said alterations may occur at the 5' or 3' terminal positions of the reference nucleotide sequence or anywhere between those terminal positions, interspersed either individually among the nucleotides in the reference sequence or in one or more contiguous groups within the reference sequence, and wherein said number of nucleotide alterations is determined by multiplying the total number of nucleotides in any of SEQ ID NO:9 to SEQ ID NO:16 by the integer defining the percent identity divided by 100

and then subtracting that product from said total number of nucleotides in any of SEQ ID NO:9 to SEQ ID NO:16, or:

$n_n \leq x_n - (x_n \bullet y)$,

wherein n_n is the number of nucleotide alterations, x_n is the total number of nucleotides in any of SEQ ID NO:9 to SEQ ID NO:16, y is 0.50 for 50%, 0.60 for 60%, 0.70 for 70%, 0.80 for 80%, 0.85 for 85%, 0.90 for 90%, 0.95 for 95%, 0.97 for 97% or 1.00 for 100%, and • is the symbol for the multiplication operator, and wherein any non-integer product of x_n and y is rounded down to the nearest integer prior to subtracting it from x_n. Alterations of polynucleotide sequences encoding the polypeptides of any of SEQ ID NO:1 to SEQ ID NO:8 may create nonsense, missense or frameshift mutations in this coding sequence and thereby alter the polypeptide encoded by the polynucleotide following such alterations.

By way of example, a polynucleotide sequence of the present invention may be identical to any of the reference sequences of SEQ ID NO:9 to SEQ ID NO:16, that is it may be 100% identical, or it may include up to a certain integer number of nucleic acid alterations as compared to the reference sequence such that the percent identity is less than 100% identity. Such alterations are selected from the group consisting of at least one nucleic acid deletion, substitution, including transition and transversion, or insertion, and wherein said alterations may occur at the 5' or 3' terminal positions of the reference polynucleotide sequence or anywhere between those terminal positions, interspersed either individually among the nucleic acids in the reference sequence or in one or more contiguous groups within the reference sequence. The number of nucleic acid alterations for a given percent identity is determined by multiplying the total number of nucleic acids in any of SEQ ID NO:9 to SEQ ID NO:16 by the integer defining the percent identity divided by 100 and then subtracting that product from said total number of nucleic acids in any of SEQ ID NO:9 to SEQ ID NO:16, or:

$n_n \leq x_n - (x_n \bullet y)$,

wherein n_n is the number of nucleic acid alterations, x_n is the total number of nucleic acids in any of SEQ ID NO:9 to SEQ ID NO:16, y is, for instance 0.70 for 70%, 0.80 for 80%, 0.85 for 85% etc., • is the symbol for the multiplication operator, and wherein any non-integer product of x_n and y is rounded down to the nearest integer prior to subtracting it from x_n.

(2) Polypeptide embodiments further include an isolated polypeptide comprising a polypeptide having at least a 50,60, 70, 80, 85, 90, 95, 97 or 100% identity to the polypeptide reference sequence of any of SEQ ID NO:1 to SEQ ID NO:8, wherein said polypeptide sequence may be identical to any of the reference sequence of SEQ ID NO:1 to SEQ ID NO:8 or may include up to a certain integer number of amino acid alterations as compared to the reference sequence, wherein said alterations are selected from the group consisting of at least one amino acid deletion, substitution, including conservative and non-conservative substitution, or insertion, and wherein said alterations may occur at the amino- or carboxy-terminal positions of the reference polypeptide sequence or anywhere between those terminal positions, interspersed either individually among the amino acids in the reference sequence or in one or more contiguous groups within the reference sequence, and wherein said number of amino acid alterations is determined by multiplying the total number of amino acids in any of SEQ ID NO:1 to SEQ ID NO:8by the integer defining the percent identity divided by 100 and then subtracting that product from said total number of amino acids in any of SEQ ID NO:1 to SEQ ID NO:8, or:

$$n_a \leq x_a - (x_a \bullet y),$$

wherein n_a is the number of amino acid alterations, x_a is the total number of amino acids in SEQ ID NO:2, y is 0.50 for 50%, 0.60 for 60%, 0.70 for 70%, 0.80 for 80%, 0.85 for 85%, 0.90 for 90%, 0.95 for 95%, 0.97 for 97% or 1.00 for 100%, and • is the symbol for the multiplication operator, and wherein any non-integer product of x_a and y is rounded down to the nearest integer prior to subtracting it from x_a.

By way of example, a polypeptide sequence of the present invention may be identical to the reference sequence of any of SEQ ID NO:1 to SEQ ID NO:8, that is it may be 100% identical, or it may include up to a certain integer number of amino acid alterations as

compared to the reference sequence such that the percent identity is less than 100% identity. Such alterations are selected from the group consisting of at least one amino acid deletion, substitution, including conservative and non-conservative substitution, or insertion, and wherein said alterations may occur at the amino- or carboxy-terminal positions of the reference polypeptide sequence or anywhere between those terminal positions, interspersed either individually among the amino acids in the reference sequence or in one or more contiguous groups within the reference sequence. The number of amino acid alterations for a given % identity is determined by multiplying the total number of amino acids in any of SEQ ID NO:1 to SEQ ID NO:8 by the integer defining the percent identity divided by 100 and then subtracting that product from said total number of amino acids in any of SEQ ID NO:1 to SEQ ID NO:8, or:

$$n_a \leq x_a - (x_a \bullet y),$$

wherein n_a is the number of amino acid alterations, x_a is the total number of amino acids in any of SEQ ID NO:1 to SEQ ID NO:8, y is, for instance 0.70 for 70%, 0.80 for 80%, 0.85 for 85% etc., and • is the symbol for the multiplication operator, and wherein any non-integer product of x_a and y is rounded down to the nearest integer prior to subtracting it from x_a.

The invention will be further described by reference to the following examples:

EXAMPLE I: Preparation of the recombinant Yeast strain Y1796 expressing P501 Fusion Protein containing a C-LytA-P2-C-LytA (CPC) as fusion partner

1. – Protein design

The structure of the fusion protein C-P2-C-p501 (alternatively named CPC-P501) to be expressed in *S. cerevisiae* is depicted in figure 2. This fusion contains the C-terminal region of gene LytA (residues 187 to 306), in which the P2 fragment of tetanus toxin (residues 830-843) has been inserted. The P2 fragment is placed between the residues 278 and 279 of C-Lyt-A. The C-lytA fragment containing the P2 insertion is followed by P501 (residues amino acid 51 to 553) and by the His tail.

The primary structure of the resulting fusion protein has the sequence described in figure 3.

The coding sequence corresponding to the above protein design is shown in figure 4.

2. – Cloning strategy for the generation of a yeast plasmid expressing CPC-P501 (51-553)-His fusion protein

- The starting material is the yeast vector pRIT15068 (UK patent application 0015619.0).
- This vector contains the yeast Cup1 promoter, the yeast alpha prepro signal coding sequence and the coding sequence corresponding to residues 55 to 553 of P501S followed by His tail.
- The cloning strategy outlined in figure 5 include the following steps:

a) The first step is the insertion of P2 sequence in frame, inside the C-lytA coding sequence. The C-lytA coding sequence is harbored by plasmid pRIT 14662 (PCT/EP99/00660). The insertion is done using an adaptor formed by two complementary oligonucleotides named P21 and P22 into the plasmid pRIT 14662 previously open by NcoI

The sequence of P21 and P22 is:

P21 5′ catgcaatacatcaaggctaactctaagttcattggtatcactgaaggcgt 3′

P22 3′ gttatgtagttccgattgagattcaagtaaccatagtgacttccgcagtac 5′

After ligation and transformation of *E. coli* and transformant characterization, the plasmid named pRIT15199 is obtained.

b) The second step is the preparation of C-lytA-P2-C-lytA DNA fragment by PCR amplification. The amplification is performed using pRIT15199 as template and the oligonucleotides named C-LytANOTATG and C-LytA-aa55. The sequence of both oligonucleotides being:

C-LytANOTATG

=5′aaaaccatggcggccgcttacgtacattccgacggctcttatccaaaagacaag 3′

C-LytA-aa55 =5' aaacatgtacatgaacttttctggcctgtctgccagtgttc 3'

The amplified fragment is treated with the restriction enzymes NcoI and Afl III to generate the respective cohesive ends.

c) The next step is the ligation of the above fragment with vector pRIT15068 (largest fragment obtained after NcoI treatment) to generate the complete fusion protein coding sequence. After ligation and *E. coli* transformation the plasmid named pRIT15200 is obtained. In this plasmid the remaining unique Nco1 site contains the ATG coding for the start codon.

d) In the next step a NcoI fragment containing the CUP1 promoter and a portion of 2μ plasmid sequences is prepared from plasmid PRIT 15202. Plasmid pRIT 15202 is a yeast 2μ derivative containing the CUP1 promoter with an NcoI site at ATG (ATG sequence: AAACC ATG)

e) The NcoI fragment isolated from pRIT 15202 is ligated to pRIT15200, previously open with NcoI, in the righ orientation, in such a way the pCUP1 promoter is at the 5' side of the coding sequence. This results in the generation of a final expression plasmid named pRIT15201(see figure 6).

3. – Preparation of the recombinant yeast strain Y1796 (RIX4440)

The plasmid pRIT 15201 is used to transform the *S. cerevisiae* strain DC5 (ATCC 20820). After selection and characterisation of the yeast transformants containing the plasmid pRIT 15201 a recombinant yeast strain named Y1796 expressing CPC-P501-His fusion protein is obtained. The protein after reduction and carboxyamidation, is isolated and purified by affinity chromatography (IMAC) followed by anion exchange chromatography (Q Sepharose FF).

Example II

In analogous fashion proteins constructs as depicted in figure 7 may be expressed utilising the corresponding DNA sequences shown therein. In particular, yeast strain SC333 (construct 2) corresponds to Y1796 strain but expressing $P501_{55-553}$ devoid of the CPC

fusion partner. Yeast strain Y1800 (construct 3) corresponds to Y1796 strain but additionally comprises the native sequence signal for P501S (aa1-aa34), while yeast strain Y1802 (construct 4) comprises the alpha pre signal sequence upstream P501S sequence. Yeast strain Y1790 (construct 5) is expressing a P501S construct devoid of CPC and having the alpha prepro signal sequence.

Example III. Preparation of purified CPC-P501

1. – Production of CPC-P501S HIS (Y1796) at small scale

For Y1796, in minimal medium supplemented with histidine, expression is induced in log phase by addition of CuSO4 ranging from 100 to 500 μM, and culture is maintained at 30°. Cells are harvested after 8 or 24H induction. Copper is added just before use and not mixed with medium in advance.

For SDS PAGE analysis, yeast cells extraction is performed in citrate phosphate buffer pH4.0 + 130 mM NaCl. Extraction is performed with glass beads for small cell quantity and with French press for higher cells quantity, and then mixed with sample buffer and SDS-PAGE analyzed.

As shown in Table 1 below, the level of expression of the culture is much higher for Y1796 strain as compared to the expression level of parent strain SC333, a strain expressing the corresponding P501S-His devoid of CPC partner. Likewise, the presence of a signal sequence (alpha pre) does not affect the results discussed above: the level of expression of the culture is much higher for Y1802 strain as compared to the expression level of corresponding strain Y1790, a strain expressing the corresponding P501S-His devoid of CPC partner.

Recombinant Strain	Plasmid	Promotor	Signal sequence	Fusion Partner	P501 aa sequences	Expression level
SC333	Ma333	CUP1	–	–	55-553-His	⊖ND
Y1796	pRIT 15201	CUP 1	–	CPC	51-553- His	+++

Y1802	TCMP28	CUP 1	α pre	CPC	51-553- His	++++
Y1790	pRIT 15068	CUP 1	α prepro	-	55-553- His	+

CPC = clyta P2 clyta

ND= not detectable

2. – Fermentation of Y1796 (RIX4440) at larger scale

100µl of the working seed are spread on solid medium and grown for approximately 24h at 30°C. This solid pre-culture is then used to inoculate a liquid pre-culture in shake flasks. This liquid pre-culture is grown for 20h at 30°C and transferred into a 20L fermenter. The fed-batch fermentation includes a growth phase of about 44h and an induction phase of about 22h.

The carbon source (glucose) was supplemented to the culture by a continuous feeding. The residual glucose concentration was maintained very low (≤50mg/L) in order to minimise the ethanol production by fermentation. This was realised by limiting the development of the micro-organism by limited glucose feed rate.
At the end of the growth phase, CUP1 promoter is induced by adding CuSO4 in order to produce the antigen.

The absence of contaminations was checked by inoculating 10^6 cells into standard TSB and THI vials supplemented with nystatine and incubated respectively for 14 days at 20-25°C and at 30-35°C. No growth was observed as expected.

3. – <u>Antigen characterisation and productivity</u>

Cell homogenates were prepared by French pressing of fermentation samples harvested at different times during the induction phase and analysed by SDS-PAGE and Western Blot. It was shown that the major part of the protein of interest was located in the insoluble fraction obtained from the cell homogenate after centrifugation. The SDS-PAGE and Western Blot analyses shown in the Figures below were realised on the pellets obtained after centrifugation of these cell homogenates.

Figures 8 A and B show a kinetics of the antigen production during the induction phase for culture PRO127. It appears that no antigen expression occurred during the growth phase. The specific antigen productivity seems to increase from the beginning of the induction phase up to 6h and then remained quite stable up to the end. But the volumetric productivity increased by a factor 1.5 to 2 due to biomass accumulation observed during the same period of time. The antigen productivity was estimated at about 500 mg per litre of fermentation broth by comparing purified reference of the antigen and crude extracts on SDS-PAGE with silver staining (figure 8A) and WB analyses using an anti-P501S antibody (a murine ascite directed against P501S aa439-aa459 used at a dilution of 1/1000) (figure 8B).

Example IV. Purification of CPC-P501 (51-553)-His fusion protein produced by Y1796

After the cell breakage, the protein is associated with the pellet fraction. A carbamidomethylation of the molecule has been introduced in the process in order to cope with the oxidative aggregation of the molecule with itself or with host cell protein contaminants through covalent bridging with disulphide bonds. The use of detergents has also been required to manage the hydrophobic character of this protein (12 trans-membrane domains predicted).

The purification protocol, developed for the scale of 1 L of culture OD (optical density) 120, is described in figure 9. All the operations are performed at room temperature (RT).

According to DOC TCA BCA protein assay, the global purification yield is 30 - 70 mg of purified antigen / L of culture OD 120. The yield is linked to the level of expression of the culture and is higher as compared to the purification yield of parent strain expressing unfused P501S-His.

The protein assay is performed as followed: proteins are first precipitated using TCA (trichloroacetic acid) in the presence of DOC (deoxycholate) then dissolved in a alcaline medium in the presence of SDS. The proteins then react with BCA (bicinchoninic acid) (Pierce) to form a soluble purple complex presenting a high adsorbance at 562 nm, which is proportional to the amount of proteins present in the sample.

SDS-PAGE analysis of 3 purified bulks (figure 10) shows no difference in reducing and non reducing conditions (cf. lanes 2, 3 and 4 versus lanes 5, 6 and 7). The pattern consists

of a major band at 70 kDa, a smear of higher MW and faint degradation bands. All the bands are detected by a specific anti P501S monoclonal antibody.

Example V. Vaccine preparation using CPC- P501S His protein

The protein of Example 3 or 4 can be formulated into a vaccine containing QS21 and 3D-MPL in an oil in water emulsion.

1. – Vaccine preparation:

The antigen produced as shown in Example 1 to 3 a C-LytA – P2 – P501S His. As an adjuvant, the formulation comprises a mixture of 3 de -O-acylated monophosphoryl lipid A (3D-MPL) and QS21 in an oil/water emulsion. The adjuvant system SBAS2 has been previously described WO 95/17210.

3D-MPL: is an immunostimulant derived from the lipopolysaccharide (LPS) of the Gram-negative bacterium *Salmonella minnesota*. MPL has been deacylated and is lacking a phosphate group on the lipid A moiety. This chemical treatment dramatically reduces toxicity while preserving the immunostimulant properties (Ribi, 1986). Ribi Immunochemistry produces and supplies MPL to SB-Biologicals.
Experiments performed at Smith Kline Beecham Biologicals have shown that
3D-MPL combined with various vehicles strongly enhances both the humoral and a TH1 type of cellular immunity.

QS21: is a natural saponin molecule extracted from the bark of the South American tree Quillaja saponaria Molina. A purification technique developed to separate the individual saponins from the crude extracts of the bark, permitted the isolation of the particular saponin, QS21, which is a triterpene glycoside demonstrating stronger adjuvant activity and lower toxicity as compared with the parent component. QS21 has been shown to activate MHC class I restricted CTLs to several subunit Ags, as well as to stimulate Ag specific lymphocytic proliferation (Kensil, 1992). Aquila (formally Cambridge Biotech Corporation) produces and supplies QS21 to SB-Biologicals.

Experiments performed at SmithKline Beecham Biologicals have demonstrated a clear synergistic effect of combinations of MPL and QS21 in the induction of both humoral and TH1 type cellular immune responses.

The oil/water emulsion is composed an organic phase made of of 2 oils (a tocopherol and squalene), and an aqueous phase of PBS containing Tween 80 as emulsifier. The emulsion comprised 5% squalene 5% tocopherol 0.4% Tween 80 and had an average particle size of 180 nm and is known as SB62 (see WO 95/17210).

Experiments performed at SmithKline Beecham Biologicals have proven that the adjunction of this O/W emulsion to 3D-MPL/QS21 (SBAS2) further increases the immunostimulant properties of the latter against various subunit antigens.

2. – Preparation of emulsion SB62 (2 fold concentrate):

Tween 80 is dissolved in phosphate buffered saline (PBS) to give a 2% solution in the PBS. To provide 100 ml two fold concentrate emulsion 5g of DL alpha tocopherol and 5ml of squalene are vortexed to mix thoroughly. 90ml of PBS/Tween solution is added and mixed thoroughly. The resulting emulsion is then passed through a syringe and finally microfluidised by using an M110S microfluidics machine. The resulting oil droplets have a size of approximately 180 nm.

3. – Formulations:

The formulations containing 3D-MPL and QS21 in an oil/water emulsion were performed as follows: 20μg – 25 ug C-LytA P2-P501S are diluted in 10 fold concentrated of PBS pH 6.8 and H2O before consecutive addition of SB62 (50μl), MPL (20μg), QS21 (20μg) and 1 μg/ml thiomersal as preservative of 5 min intervals. All incubations are carried out at room temperature with agitation.

Example VI. Immunogenicity experiments

1. – Mice studies

The immune response induced by vaccination using the recombinant purified CPCP501S protein formulated in adjuvants is characterised in experiments performed in mice.

Groups of 5 to 10, eight weeks old female mice (C57BL6 or CB6F1 hybrid of C57BL/6 and Balb/C mice) are vaccinated, 2- 4 times intra-muscularly at 2 weeks intervals with 10 μg of the CPCP501S protein formulated in different adjuvant systems. The volume administered corresponds to $1/10^{th}$ of a human dose (50 μl).

The serology (total IgG response and isotypic profile) and cellular response (T cell lymphoproliferation, cytolytic activity and cytokine production) are analysed on spleen or lymph node cells, 14 days after 2 or 4 vaccinations using standard protocols as described in Gérard, c. et al, 2001, Vaccine 19, 2583-2589.

Claims

1. A fusion partner protein comprising a choline binding domain and a heterologous promiscuous T helper epitope.
2. A fusion partner protein according to claim 1 wherein the choline binding domain is selected from the group comprising:
   a) the C-terminal domain of LytA as set forth in SEQ ID NO:7; or
   b) the sequence of SEQ ID NO:8; or
   c) a peptide sequence comprising an amino acid sequence having at least 85% identity, preferably at least 90% identity, more preferably at least 95% identity, most preferably at least 97-99% identity, to any of SEQ ID NO:1 to 6; or
   d) a peptide sequence comprising an amino acid sequence having at least 15, 20, 30, 40, 50 or 100 contiguous amino acids from the amino acid sequence of SEQ ID NO:7 or SEQ ID NO:8.
3. A fusion partner protein as claimed in claim 1 or 2 further comprising a heterologous protein.
4. A fusion protein as claimed in claim 3 wherein the heterologous protein is chemically conjugated the fusion partner.
5. A fusion protein as claimed in claim 3 or 4 wherein the heterologous protein is a tumour associated protein or tissue specific protein or immunogenic fragment thereof.
6. A fusion protein as claimed in any of claims 3 to 5 wherein the heterologous protein or fragment thereof is selected from MAGE 1, MAGE 3, MAGE 4, PRAME, BAGE, LAGE, SAGE, HAGE, PSA, PAP, PSCA, prostein, HASH2, Cripto, Prostase, STEAP, tyrosinase, telomerase, survivin, or her 2 neu.
7. A fusion protein as claimed in any of claims 4 to 6 further comprising an affinity tag of at least 4 histidine residues.
8. A nucleic acid sequence encoding a protein of claim 1 to 7.
9. An expression vector comprising a nucleic acid sequence of claim 8.
10. A host transformed with a nucleic acid sequence of claim 8 or with an expression vector of claim 9.

11. An immunogenic composition comprising a protein as claimed in any of claim 1 to 7 or a DNA sequence as claimed in claim 8 and a pharmaceutically acceptable excipient.

12. An immunogenic composition as claimed in claim 11 which additionally comprises a TH-1 inducing adjuvant.

13. An immunogenic composition as claimed in claim 12 in which the TH-1 inducing adjuvant is selected from the group of adjuvants comprising: 3D-MPL, QS21, a mixture of QS21 and cholesterol, a CpG oligonucleotide or a mixture of two or more said adjuvants.

14. A process for the preparation of a immunogenic composition as claimed in any of claims 11 to 13, comprising admixing the fusion protein of any of claims 4 to 7 or a the encoding polynucleotide of claim 8 with a suitable adjuvant, diluent or other pharmaceutically acceptable carrier.

15. A process for producing a fusion protein of any of claims 1 to 7 comprising culturing a host cell of claim 10 under conditions sufficient for the production of said fusion protein and recovering the fusion protein from the culture medium.

16. A protein of any of claims 1 to 7 or a DNA sequence of claim 8 for use in medicine.

17. Use of a protein as claimed in any of claim 1 to 7 or a DNA sequence of claim 8 in the manufacture of a immunogenic composition for immunotherapeutically treating a patient suffering from or susceptible to cancer.

18. A method of treating a patient suffering from cancer by administrating a safe and effective amount of a composition of claim 9.

Figure 1 – Sequence information for C-LytA.

Each repeat has been defined on the basis of both multiple sequence alignment and secondary structure prediction using the following alignment programs:

1) MatchBox (Depiereux E et al. (1992) Comput Applic Biosci 8:501-9)
2) ClustalW (Thompson JD et al. (1994) Nucl Acid Res 22:4673-80)
3) Block-Maker (Henikoff S et al (1995) Gene 163:gc17-26)

SEQ ID NO:1 – amino acid sequence of C-LytA repeat 1

GWQKNDTGYWYVHSD 15

SEQ ID NO:2 – amino acid sequence of C-LytA repeat 2

GSYPKDKFEKINGTWYYFDSS 21

SEQ ID NO:3 – amino acid sequence of C-LytA repeat 3

GYMLADRWRKHTDGNWYWFDNS 22

SEQ ID NO:4 – amino acid sequence of C-LytA repeat 4

GEMATGWKKIADKWYYFNEE 20

SEQ ID NO:5 – amino acid sequence of C-LytA repeat 5

GAMKTGWVKYKDTWYYLDAKE 21

SEQ ID NO:6 – amino acid sequence of C-LytA repeat 6

GAMVSNAFIQSADGTGWYYLKPD 23

SEQ ID NO:7 – amino acid sequence of C-LytA cholin-binding domain

GWQKNDTGYW YVHSDGSYPK DKFEKINGTW YYFDSSGYML ADRWRKHTDG NWYWFDNSGE 60
MATGWKKIAD KWYYFNEEGA MKTGWVKYKD TWYYLDAKEG AMVSNAFIQS ADGTGWYYLK 120
PDGTLADRPE FTVEPDGLIT VK 142

SEQ ID NO:8 – amino acid sequence of C-LytA domain from truncated repeat 1 to repeat 6 (as part of our constructs shown in figure 7)

YVHSDGSYPKDKFEKINGTWYYFDSSGYMLADRWRKHTDGNWYWFDNSGEMATGWKKIADKWYYFNEEGAMKT
GWVKYKDTWYYLDAKEGAMVSNAFIQSADGTGWYYLKPD

SEQ ID NO:9 – DNA sequence encoding the amino acid sequence of SEQ ID NO:1

ggctggcaga agaatgacac tggctactgg tacgtacatt cagac

SEQ ID NO:10 – DNA sequence encoding the amino acid sequence of SEQ ID NO:2

ggctcttatc caaaagacaa gtttgagaaa atcaatggca cttggtacta ctttgacagt tca

SEQ ID NO:11 – DNA sequence encoding the amino acid sequence of SEQ ID NO:3

ggctatatgc ttgcagaccg ctggaggaag cacacagacg gcaactggta ctggttcgac aactca

SEQ ID NO:12 – DNA sequence encoding the amino acid sequence of SEQ ID NO:4

ggcgaaatgg ctacaggctg gaagaaaatc gctgataagt ggtactattt caacgaagaa

SEQ ID NO:13 – DNA sequence encoding the amino acid sequence of SEQ ID NO:5

Ggtgccatga agacaggctg ggtcaagtac aaggacactt ggtactactt agacgctaaa gaa

SEQ ID NO:14 – DNA sequence encoding the amino acid sequence of SEQ ID NO:6

Ggcgccatgg tatcaaatgc ctttatccag tcagcggacg gaacaggctg gtactacctc
aaaccagac

SEQ ID NO:15 – DNA sequence encoding the amino acid sequence of SEQ ID NO:7

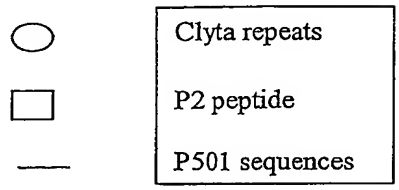
ggctggcaga agaatgacac tggctactgg tacgtacatt cagacggctc ttatccaaaa 60
gacaagtttg agaaaatcaa tggcacttgg tactactttg acagttcagg ctatatgctt 120
gcagaccgct ggaggaagca cacagacggc aactggtact ggttcgacaa ctcaggcgaa 180
atggctacag gctggaagaa aatcgctgat aagtggtact atttcaacga agaaggtgcc 240
atgaagacag gctgggtcaa gtacaaggac acttggtact acttagacgc taaagaaggc 300
gccatggtat caaatgcctt tatccagtca gcggacggaa caggctggta ctacctcaaa 360
ccagacggaa cactggcaga caggccagaa ttcacagtag agccagatgg cttgattaca 420
gtaaaataa 429

SEQ ID NO:16 – DNA sequence encoding the amino acid sequence of SEQ ID NO:8

B45311

TACGTACATTCCGACGGCTCTTATCCAAAAGACAAGTTTGAGAAAATCAATGGCACTTGGTACTACTTTGACA
GTTCAGGCTATATGCTTGCAGACCGCTGGAGGAAGCACACAGACGGCAACTGGTACTGGTTCGACAACTCAGG
CGAAATGGCTACAGGCTGGAAGAAAATCGCTGATAAGTGGTACTATTTCAACGAAGAAGGTGCCATGAAGACA
GGCTGGGTCAAGTACAAGGACACTTGGTACTACTTAGACGCTAAAGAAGGCGCCATGGTATCAAATGCCTTTA
TCCAGTCAGCGGACGGAACAGGCTGGTACTACCTCAAACCAGAC

Figure 2. Structure of CPC-p501 His fusion protein expressed in *S. cerevisiae*



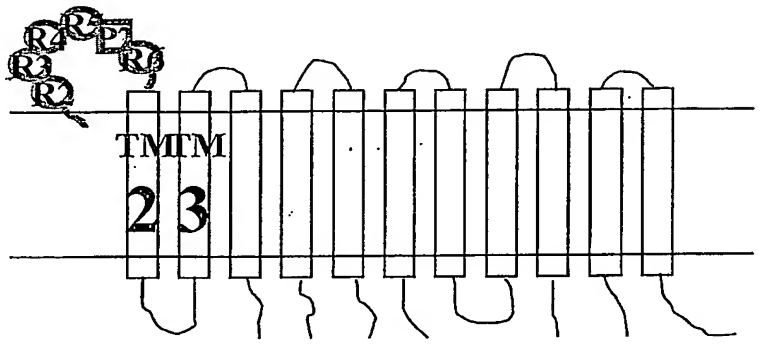


Figure 3. Primary structure of CPC-P501 His fusion protein

```
MAAAYVHSDG SYPKDKFEKI NGTWYYFDSS GYMLADRWRK HTDGNWYWFD NSGEMATGWK  60
KIADKWYYFN EEGAMKTGWV KYKDTWYYLD AKEGAMQYIK ANSKFIGITE GVMVSNAFIQ  120
SADGTGWYYL KPDGTLADRP EKFMYMVLGI GPVLGLVCVP LLGSASDHWR GRYGRRRPFI  180
WALSLGILLS LFLIPRAGWL AGLLCPDPRP LELALLILGV GLLDFCGQVC FTPLEALLSD  240
LFRDPDHCRQ AYSVYAFMIS LGGCLGYLLP AIDWDTSALA PYLGTQEECL FGLLTLIFLT  300
CVAATLLVAE EAALGPTEPA EGLSAPSLSP HCCPCRARLA FRNLGALLPR LHQLCCRMPR  360
TLRRLFVAEL CSWMALMTFT LFYTDFVGEG LYQGVPRAEP GTEARRHYDE GVRMGSLGLF  420
LQCAISLVFS LVMDRLVQRF GTRAVYLASV AAFPVAAGAT CLSHSVAVVT ASAALTGFTF  480
SALQILPYTL ASLYHREKQV FLPKYRGDTG GASSEDSLMT SFLPGPKPGA PFPNGHVGAG  540
GSGLLPPPPA LCGASACDVS VRVVVGEPTE ARVVPGRGIC LDLAILDSAF LLSQVAPSLF  600
MGSIVQLSQS VTAYMVSAAG LGLVAIYFAT QVVFDKSDLA KYSAGGHHHH HH  652
```

Figure 4. Nucleotide sequence of CPC P501 His(pRIT15201)

```
ATGGCGGCCG CTTACGTACA TTCCGACGGC TCTTATCCAA AAGACAAGTT TGAGAAAATC   60
AATGGCACTT GGTACTACTT TGACAGTTCA GGCTATATGC TTGCAGACCG CTGGAGGAAG  120
CACACAGACG GCAACTGGTA CTGGTTCGAC AACTCAGGCG AAATGGCTAC AGGCTGGAAG  180
AAAATCGCTG ATAAGTGGTA CTATTTCAAC GAAGAAGGTG CCATGAAGAC AGGCTGGGTC  240
AAGTACAAGG ACACTTGGTA CTACTTAGAC GCTAAAGAAG GCGCCATGCA ATACATCAAG  300
GCTAACTCTA AGTTCATTGG TATCACTGAA GGCGTCATGG TATCAAATGC CTTTATCCAG  360
TCAGCGGACG GAACAGGCTG GTACTACCTC AAACCAGACG GAACACTGGC AGACAGGCCA  420
GAAAAGTTCA TGTACATGGT GCTGGGCATT GGTCCAGTGC TGGGCCTGGT CTGTGTCCCG  480
CTCCTAGGCT CAGCCAGTGA CCACTGGCGT GGACGCTATG GCCGCCGCCG GCCCTTCATC  540
TGGGCACTGT CCTTGGGCAT CCTGCTGAGC CTCTTTCTCA TCCCAAGGGC CGGCTGGCTA  600
GCAGGGCTGC TGTGCCCGGA TCCCAGGCCC CTGGAGCTGG CACTGCTCAT CCTGGGCGTG  660
GGGCTGCTGG ACTTCTGTGG CCAGGTGTGC TTCACTCCAC TGGAGGCCCT GCTCTCTGAC  720
CTCTTCCGGG ACCCGGACCA CTGTCGCCAG GCCTACTCTG TCTATGCCTT CATGATCAGT  780
CTTGGGGGCT GCCTGGGCTA CCTCCTGCCT GCCATTGACT GGGACACCAG TGCCCTGGCC  840
CCCTACCTGG GCACCCAGGA GGAGTGCCTC TTTGGCCTGC TCACCCTCAT CTTCCTCACC  900
TGCGTAGCAG CCACACTGCT GGTGGCTGAG GAGGCAGCGC TGGGCCCCAC CGAGCCAGCA  960
GAAGGGCTGT CGGCCCCCTC CTTGTCGCCC CACTGCTGTC CATGCCGGGC CCGCTTGGCT 1020
TTCCGGAACC TGGGCGCCCT GCTTCCCCGG CTGCACCAGC TGTGCTGCCG CATGCCCCGC 1080
ACCCTGCGCC GGCTCTTCGT GGCTGAGCTG TGCAGCTGGA TGGCACTCAT GACCTTCACG 1140
CTGTTTTACA CGGATTTCGT GGGCGAGGGG CTGTACCAGG GCGTGCCCAG AGCTGAGCCG 1200
GGCACCGAGG CCCGGAGACA CTATGATGAA GGCGTTCGGA TGGGCAGCCT GGGGCTGTTC 1260
CTGCAGTGCG CCATCTCCCT GGTCTTCTCT CTGGTCATGG ACCGGCTGGT GCAGCGATTC 1320
GGCACTCGAG CAGTCTATTT GGCCAGTGTG GCAGCTTTCC CTGTGGCTGC CGGTGCCACA 1380
TGCCTGTCCC ACAGTGTGGC CGTGGTGACA GCTTCAGCCG CCCTCACCGG GTTCACCTTC 1440
TCAGCCCTGC AGATCCTGCC CTACACACTG GCCTCCCTCT ACCACCGGGA GAAGCAGGTG 1500
TTCCTGCCCA AATACCGAGG GGACACTGGA GGTGCTAGCA GTGAGGACAG CCTGATGACC 1560
AGCTTCCTGC CAGGCCCTAA GCCTGGAGCT CCCTTCCCTA ATGGACACGT GGGTGCTGGA 1620
GGCAGTGGCC TGCTCCCACC TCCACCCGCG CTCTGCGGGG CCTCTGCCTG TGATGTCTCC 1680
GTACGTGTGG TGGTGGGTGA GCCCACCGAG GCCAGGGTGG TTCCGGGCCG GGGCATCTGC 1740
CTGGACCTCG CCATCCTGGA TAGTGCCTTC CTGCTGTCCC AGGTGGCCCC ATCCCTGTTT 1800
ATGGGCTCCA TTGTCCAGCT CAGCCAGTCT GTCACTGCCT ATATGGTGTC TGCCGCAGGC 1860
CTGGGTCTGG TCGCCATTTA CTTTGCTACA CAGGTAGTAT TTGACAAGAG CGACTTGGCC 1920
AAATACTCAG CGGGTGGACA CCATCACCAT CACCATTAA  1959
```

Figure 5. Cloning strategy for generation of plasmid pRIT 15201

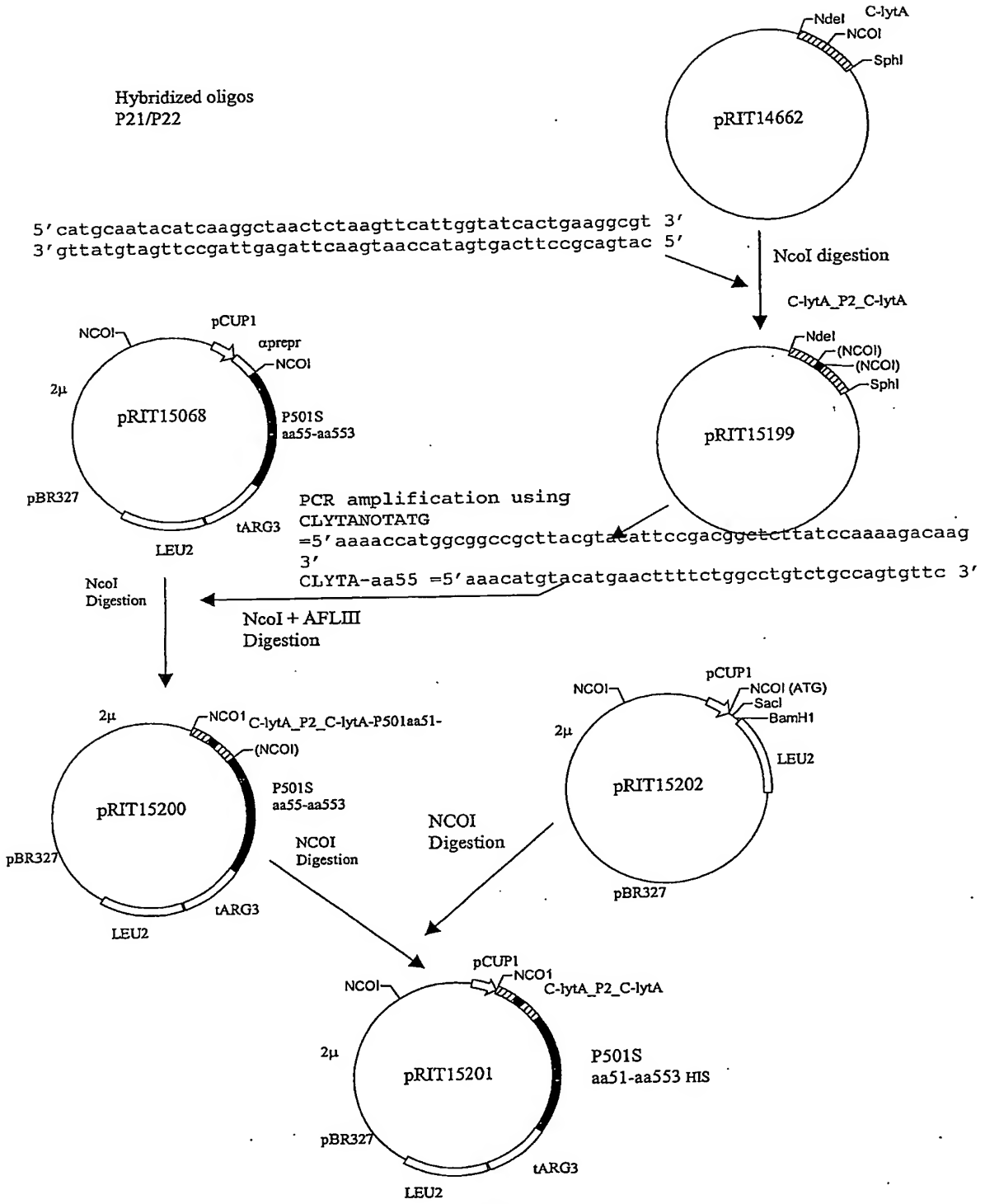


Figure 6. Plasmid map of pRIT15201

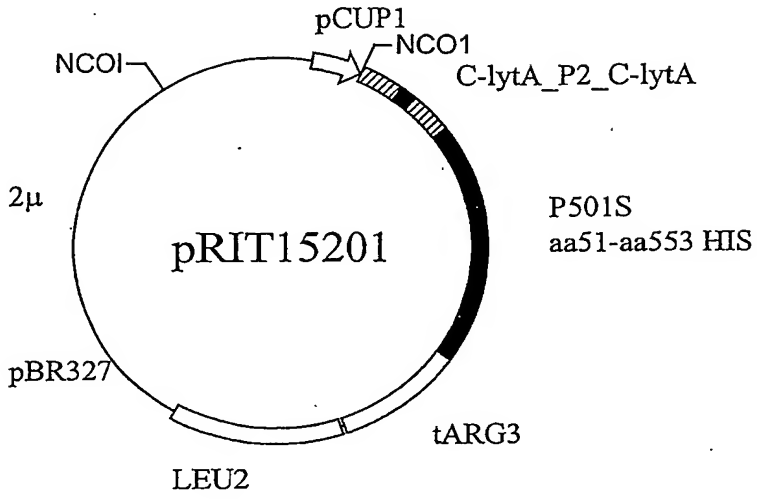


Figure 7. CPC and native Constructs

Construct 1 – coding sequence of CPC-P501$_{51-553}$ (see plasmid of figure 6 -Y1796)

Protein sequence

R1 R2 R3 R4

MAAAYVHSDGSYPKDKFEKINGTWYYFDSSGYMLADRWRKHTDGNWYWFDNSGEMATG

R5 P2 R6

WKKIADKWYYFNEEGAMKTGWVKYKDTWYYLDAKEGAMQYIKANSKFIGITEGVMVSNAFIQS

ADGTGWYYLKPDGTLADRPEKFMYMVLGIGPVLGLVCVPLLGSASDHWRGRYGRRRPFIWALSL

GILLSLFLIPRAGWLAGLLCPDPRPLELALLILGVGLLDFCGQVCFTPLEALLSDLFRDPDHCRQAYSV

YAFMISLGGCLGYLLPAIDWDTSALAPYLGTQEECLFGLLTLIFLTCVAATLLVAEEAALGPTEPAEG

LSAPSLSPHCCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFVGE

GLYQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTRAVYLASVAAFPV

AAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDTGGASSEDSLMTSF

LPGPKPGAPFPNGHVGAGGSGLLPPPPALCGASACDVSVRVVVGEPTEARVVPGRGICLDLAILDSAF

LLSQVAPSLFMGSIVQLSQSVTAYMVSAAGLGLVAIYFATQVVFDKSDLAKYSAGGHHHHHH

R1 (plain): aa5-9 (fragment) **R4 (bold)**: aa53-72 P2 (underline): 97-110

R2 (bold): aa10-30 R5 (plain): aa73-92

R3 (plain): aa31-52 **R6a (bold)**: aa94-95 **R6b (bold)**: 113-133

Nucleotide sequence

ATGgcggccgcttACGTACATTCCGACGGCTCTTATCCAAAAGACAAGTTTGAGAAAATCAATGGCACTTGGT

ACTACTTTGACAGTTCAGGCTATATGCTTGCAGACCGCTGGAGGAAGCACACAGACGGCAACTGGTACTGGTT

CGACAACTCAGGCGAAATGGCTACAGGCTGGAAGAAAATCGCTGATAAGTGGTACTATTTCAACGAAGAAGGT

GCCATGAAGACAGGCTGGGTCAAGTACAAGGACACTTGGTACTACTTAGACGCTAAAGAAGGCGCCatgcaat

acatcaaggctaactctaagttcattggtatcactgaaggcgtcATGGTATCAAATGCCTTTATCCAGTCAGC

GGACGGAACAGGCTGGTACTACCTCAAACCAGACGGAACACTGGCAGACAGGCCAGAAaagttcatgtaCatg

GTGCTGGGCATTGGTCCAGTGCTGGGCCTGGTCTGTGTCCCGCTCCTAGGCTCAGCCAGTGACCACTGGCGTG

GACGCTATGGCCGCCGCCGGCCCTTCATCTGGGCACTGTCCTTGGGCATCCTGCTGAGCCTCTTTCTCATCCC

AAGGGCCGGCTGGCTAGCAGGGCTGCTGTGCCCGGATCCCAGGCCCCTGGAGCTGGCACTGCTCATCCTGGGC

GTGGGGCTGCTGGACTTCTGTGGCCAGGTGTGCTTCACTCCACTGGAGGCCCTGCTCTCTGACCTCTTCCGGG

ACCCGGACCACTGTCGCCAGGCCTACTCTGTCTATGCCTTCATGATCAGTCTTGGGGGCTGCCTGGGCTACCT

CCTGCCTGCCATTGACTGGGACACCAGTGCCCTGGCCCCCTACCTGGGCACCCAGGAGGAGTGCCTCTTTGGC

CTGCTCACCCTCATCTTCCTCACCTGCGTAGCAGCCACACTGCTGGTGGCTGAGGAGGCAGCGCTGGGCCCCA

CCGAGCCAGCAGAAGGGCTGTCGGCCCCCTCCTTGTCGCCCCACTGCTGTCCATGCCGGGCCCGCTTGGCTTT

CCGGAACCTGGGCGCCCTGCTTCCCCGGCTGCACCAGCTGTGCTGCCGCATGCCCCGCACCCTGCGCCGGCTC

TTCGTGGCTGAGCTGTGCAGCTGGATGGCACTCATGACCTTCACGCTGTTTTACACGGATTTCGTGGGCGAGG

GGCTGTACCAGGGCGTGCCCAGAGCTGAGCCGGGCACCGAGGCCCGGAGACACTATGATGAAGGCGTTCGGAT
GGGCAGCCTGGGGCTGTTCCTGCAGTGCGCCATCTCCCTGGTCTTCTCTCTGGTCATGGACCGGCTGGTGCAG
CGATTCGGCACTCGAGCAGTCTATTTGGCCAGTGTGGCAGCTTTCCCTGTGGCTGCCGGTGCCACATGCCTGT
CCCACAGTGTGGCCGTGGTGACAGCTTCAGCCGCCCTCACCGGGTTCACCTTCTCAGCCCTGCAGATCCTGCC
CTACACACTGGCCTCCCTCTACCACCGGGAGAAGCAGGTGTTCCTGCCCAAATACCGAGGGGACACTGGAGGT
GCTAGCAGTGAGGACAGCCTGATGACCAGCTTCCTGCCAGGCCCTAAGCCTGGAGCTCCCTTCCCTAATGGAC
ACGTGGGTGCTGGAGGCAGTGGCCTGCTCCCACCTCCACCCGCGCTCTGCGGGGCCTCTGCCTGTGAtGTCTC
CGTACGTGTGGTGGTGGGTGAGCCCACCGAGGCCAGGGTGGTTCCGGGCCGGGGCATCTGCCTGGACCTCGCC
ATCCTGGATAGTGCCTTCCTGCTGTCCCAGGTGGCCCCATCCCTGTTTATGGGCTCCATTGTCCAGCTCAGCC
AGTCTGTCACTGCCTATATGGTGTCTGCCGCAGGCCTGGGTCTGGTCGCCATTTACTTTGCTACACAGGTAGT
ATTTGACAAGAGCGACTTGGCCAAATACTCAGCGggtggacaccatcaccatcaccattaa

Construct 2 – Coding sequence of $P501_{55-553}$ HIS (control) (yeast strain SC333)

Protein sequence

MVLGIGPVLG LVCVPLLGSA SDHWRGRYGR RRPFIWALSL GILLSLFLIP RAGWLAGLLC 60
PDPRPLELAL LILGVGLLDF CGQVCFTPLE ALLSDLFRDP DHCRQAYSVY AFMISLGGCL 120
GYLLPAIDWD TSALAPYLGT QEECLFGLLT LIFLTCVAAT LLVAEEAALG PTEPAEGLSA 180
PSLSPHCCPC RARLAFRNLG ALLPRLHQLC CRMPRTLRRL FVAELCSWMA LMTFTLFYTD 240
FVGEGLYQGV PRAEPGTEAR RHYDEGVRMG SLGLFLQCAI SLVFSLVMDR LVQRFGTRAV 300
YLASVAAFPV AAGATCLSHS VAVVTASAAL TGFTFSALQI LPYTLASLYH REKQVFLPKY 360
RGDTGGASSE DSLMTSFLPG PKPGAPFPNG HVGAGGSGLL PPPPALCGAS ACDVSVRVVV 420
GEPTEARVVP GRGICLDLAI LDSAFLLSQV APSLFMGSIV QLSQSVTAYM VSAAGLGLVA 480
IYFATQVVFD KSDLAKYSAG GHHHHHH 507

Nucleotide sequence

atgGTGCTGG GCATTGGTCC AGTGCTGGGC CTGGTCTGTG TCCCGCTCCT AGGCTCAGCC 60
AGTGACCACT GGCGTGGACG CTATGGCCGC CGCCGGCCCT TCATCTGGGC ACTGTCCTTG 120
GGCATCCTGC TGAGCCTCTT TCTCATCCCA AGGGCCGGCT GGCTAGCAGG GCTGCTGTGC 180
CCGGATCCCA GGCCCCTGGA GCTGGCACTG CTCATCCTGG GCGTGGGGCT GCTGGACTTC 240
TGTGGCCAGG TGTGCTTCAC TCCACTGGAG GCCCTGCTCT CTGACCTCTT CCGGGACCCG 300
GACCACTGTC GCCAGGCCTA CTCTGTCTAT GCCTTCATGA TCAGTCTTGG GGGCTGCCTG 360
GGCTACCTCC TGCCTGCCAT TGACTGGGAC ACCAGTGCCC TGGCCCCCTA CCTGGGCACC 420
CAGGAGGAGT GCCTCTTTGG CCTGCTCACC CTCATCTTCC TCACCTGCGT AGCAGCCACA 480
CTGCTGGTGG CTGAGGAGGC AGCGCTGGGC CCCACCGAGC CAGCAGAAGG GCTGTCGGCC 540
CCCTCCTTGT CGCCCCACTG CTGTCCATGC CGGGCCCGCT TGGCTTTCCG GAACCTGGGC 600
GCCCTGCTTC CCCGGCTGCA CCAGCTGTGC TGCCGCATGC CCCGCACCCT GCGCCGGCTC 660
TTCGTGGCTG AGCTGTGCAG CTGGATGGCA CTCATGACCT TCACGCTGTT TTACACGGAT 720
TTCGTGGGCG AGGGGCTGTA CCAGGGCGTG CCCAGAGCTG AGCCGGGCAC CGAGGCCCGG 780
AGACACTATG ATGAAGGCGT TCGGATGGGC AGCCTGGGGC TGTTCCTGCA GTGCGCCATC 840

TCCCTGGTCT TCTCTCTGGT CATGGACCGG CTGGTGCAGC GATTCGGCAC TCGAGCAGTC 900
TATTTGGCCA GTGTGGCAGC TTTCCCTGTG GCTGCCGGTG CCACATGCCT GTCCCACAGT 960
GTGGCCGTGG TGACAGCTTC AGCCGCCCTC ACCGGGTTCA CCTTCTCAGC CCTGCAGATC 1020
CTGCCCTACA CACTGGCCTC CCTCTACCAC CGGGAGAAGC AGGTGTTCCT GCCCAAATAC 1080
CGAGGGGACA CTGGAGGTGC TAGCAGTGAG GACAGCCTGA TGACCAGCTT CCTGCCAGGC 1140
CCTAAGCCTG GAGCTCCCTT CCCTAATGGA CACGTGGGTG CTGGAGGCAG TGGCCTGCTC 1200
CCACCTCCAC CCGCGCTCTG CGGGGCCTCT GCCTGTGATG TCTCCGTACG TGTGGTGGTG 1260
GGTGAGCCCA CCGAGGCCAG GGTGGTTCCG GGCCGGGGCA TCTGCCTGGA CCTCGCCATC 1320
CTGGATAGTG CCTTCCTGCT GTCCCAGGTG GCCCCATCCC TGTTTATGGG CTCCATTGTC 1380
CAGCTCAGCC AGTCTGTCAC TGCCTATATG GTGTCTGCCG CAGGCCTGGG TCTGGTCGCC 1440
ATTTACTTTG CTACACAGGT AGTATTTGAC AAGAGCGACT TGGCCAAATA CTCAGCGggt 1500
ggacaccatc accatcacca ttaa 1524

Construct 3 - Coding sequence of natssP501$_{1-34}$ P501$_{51-553}$ HIS (yeast strain Y1800)

R1 R2

MAAVQRLWVSRLLRHRKAQLLLVNLLTFGLEVCLAAA YVHSDGSYPKDKFEKINGTW

R3 R4 R5

YYFDSSGYMLADRWRKHTDGNWYWFDNSGEMATGWKKIADKWYYFNEEGAMKTGWVK

P2. R6

YKDTWYYLDAKEGAMQYIKANSKFIGITEGV MVSNAFIQSADGTGWYYLKPD GTLADRPEKFMY
MVLGIGPVLGLVCVPLLGSASDHWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGLLCPDPRPLEL
ALLILGVGLLDFCGQVCFTPLEALLSDLFRDPDHCRQAYSVYAFMISLGGCLGYLLPAIDWDTSALAP
YLGTQEECLFGLLTLIFLTCVAATLLVAEEAALGPTEPAEGLSAPSLSPHCCPCRARLAFRNLGALLPR
LHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFVGEGLYQGVPRAEPGTEARRHYDEGVRMG
SLGLFLQCAISLVFSLVMDRLVQRFGTRAVYLASVAAFPVAAGATCLSHSVAVVTASAALTGFTFSA
LQILPYTLASLYHREKQVFLPKYRGDTGGASSEDSLMTSFLPGPKPGAPFPNGHVGAGGSGLLPPPPA
LCGASACDVSVRVVVGEPTEARVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIVQLSQSVTAYMVS
AAGLGLVAIYFATQVVFDKSDLAKYSAGGHHHHHH

R1 (plain): aa38-42 (fragment) **R4 (bold)**: aa77-106 P2 (underline): 130-143

R2 (bold): aa43-64 R5 (plain): aa107-126

R3 (plain): aa65-76 **R6a (bold)**: aa127-128 **R6b (bold)**: aa146-166

natss stands for native signal sequence

ATGgcGGCCGTGCAGAGGCTATGGGTATCGAGACTGCTAAGACACCGCAAAGCTCAGTTGTTGTTGGTTAACT
TGTTGACCTTCGGGCTGGAAGTCTGTTTGGCggccgctTACGTACATTCCGACGGCTCTTATCCAAAAGACAA
GTTTGAGAAAATCAATGGCACTTGGTACTACTTTGACAGTTCAGGCTATATGCTTGCAGACCGCTGGAGGAAG
CACACAGACGGCAACTGGTACTGGTTCGACAACTCAGGCGAAATGGCTACAGGCTGGAAGAAAATCGCTGATA
AGTGGTACTATTTCAACGAAGAAGGTGCCATGAAGACAGGCTGGGTCAAGTACAAGGACACTTGGTACTACTT

AGACGCTAAAGAAGGCGCCatgcaatacatcaaggctaactctaagttcattggtatcactgaaggcgtcATG
GTATCAAATGCCTTTATCCAGTCAGCGGACGGAACAGGCTGGTACTACCTCAAACCAGACGGAACACTGGCAG
ACAGGCCAGAAaagttcatgtaCatgGTGCTGGGCATTGGTCCAGTGCTGGGCCTGGTCTGTGTCCCGCTCCT
AGGCTCAGCCAGTGACCACTGGCGTGGACGCTATGGCCGCCGCCGGCCCTTCATCTGGGCACTGTCCTTGGGC
ATCCTGCTGAGCCTCTTTCTCATCCCAAGGGCCGGCTGGCTAGCAGGGCTGCTGTGCCCGGATCCCAGGCCCC
TGGAGCTGGCACTGCTCATCCTGGGCGTGGGGCTGCTGGACTTCTGTGGCCAGGTGTGCTTCACTCCACTGGA
GGCCCTGCTCTCTGACCTCTTCCGGGACCCGGACCACTGTCGCCAGGCCTACTCTGTCTATGCCTTCATGATC
AGTCTTGGGGGCTGCCTGGGCTACCTCCTGCCTGCCATTGACTGGGACACCAGTGCCCTGGCCCCCTACCTGG
GCACCCAGGAGGAGTGCCTCTTTGGCCTGCTCACCCTCATCTTCCTCACCTGCGTAGCAGCCACACTGCTGGT
GGCTGAGGAGGCAGCGCTGGGCCCCACCGAGCCAGCAGAAGGGCTGTCGGCCCCCTCCTTGTCGCCCCACTGC
TGTCCATGCCGGGCCCGCTTGGCTTTCCGGAACCTGGGCGCCCTGCTTCCCCGGCTGCACCAGCTGTGCTGCC
GCATGCCCCGCACCCTGCGCCGGCTCTTCGTGGCTGAGCTGTGCAGCTGGATGGCACTCATGACCTTCACGCT
GTTTTACACGGATTTCGTGGGCGAGGGGCTGTACCAGGGCGTGCCCAGAGCTGAGCCGGGCACCGAGGCCCGG
AGACACTATGATGAAGGCGTTCGGATGGGCAGCCTGGGGCTGTTCCTGCAGTGCGCCATCTCCCTGGTCTTCT
CTCTGGTCATGGACCGGCTGGTGCAGCGATTCGGCACTCGAGCAGTCTATTTGGCCAGTGTGGCAGCTTTCCC
TGTGGCTGCCGGTGCCACATGCCTGTCCCACAGTGTGGCCGTGGTGACAGCTTCAGCCGCCCTCACCGGGTTC
ACCTTCTCAGCCCTGCAGATCCTGCCCTACACACTGGCCTCCCTCTACCACCGGGAGAAGCAGGTGTTCCTGC
CCAAATACCGAGGGGACACTGGAGGTGCTAGCAGTGAGGACAGCCTGATGACCAGCTTCCTGCCAGGCCCTAA
GCCTGGAGCTCCCTTCCCTAATGGACACGTGGGTGCTGGAGGCAGTGGCCTGCTCCCACCTCCACCCGCGCTC
TGCGGGGCCTCTGCCTGTGATGTCTCCGTACGTGTGGTGGTGGGTGAGCCCACCGAGGCCAGGGTGGTTCCGG
GCCGGGGCATCTGCCTGGACCTCGCCATCCTGGATAGTGCCTTCCTGCTGTCCCAGGTGGCCCCATCCCTGTT
TATGGGCTCCATTGTCCAGCTCAGCCAGTCTGTCACTGCCTATATGGTGTCTGCCGCAGGCCTGGGTCTGGTC
GCCATTTACTTTGCTACACAGGTAGTATTTGACAAGAGCGACTTGGCCAAATACTCAGCGggtggacaccatc
accatcaccattaa

Construct 4 - Coding sequence of alphapreCPC-P501$_{51-553}$ HIS (yeast strain Y1802)

Protein sequence

Alpha-pre signal R1 R2 R3

MAARFPSIFTAVLFAASSALAAAYVHSDGSYPKDKFEKINGTWYYFDSSGYMLADRWRKHTDGNWYWFD

R4 R5 P2

NSGEMATGWKKIADKWYYFNEEGAMKTGWVKYKDTWYYLDAKEGAMQYIKANSKFIGITEGVMVSNAFI

R6

QSADGTGWYYLKPDGTLADRPEKFMYMVLGIGPVLGLVCVPLLGSASDHWRGRYGRRRPFIWALSLGILLSLF
LIPRAGWLAGLLCPDPRPLELALLILGVGLLDFCGQVCFTPLEALLSDLFRDPDHCRQAYSVYAFMISLGGCL
GYLLPAIDWDTSALAPYLGTQEECLFGLLTLIFLTCVAATLLVAEEAALGPTEPAEGLSAPSLSPHCCPCRAR
LAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFVGEGLYQGVPRAEPGTEARRHYDEG
VRMGSLGLFLQCAISLVFSLVMDRLVQRFGTRAVYLASVAAFPVAAGATCLSHSVAVVTASAALTGFTFSALQ
ILPYTLASLYHREKQVFLPKYRGDTGGASSEDSLMTSFLPGPKPGAPFPNGHVGAGGSGLLPPPPALCGASAC
DVSVRVVVGEPTEARVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIVQLSQSVTAYMVSAAGLGLVAIYFAT
QVVFDKSDLAKYSAGGHHHHHH

Alpha-pre signal (bold): aa4-22

R1 (plain): aa24-28 (fragment) **R4 (bold)**: aa72-91 P2 (underline): 116-129

R2 (bold): aa29-49 R5 (plain): aa92-112

R3 (plain): aa50-71 **R6a (bold)**: aa113-114 **R6b (bold)**: aa132-152

Alphapre stands for alpha pre signal sequence

Nucleotide sequence

TACGTACATTCCGACGGCTCTTATCCAAAAGACAAGTTTGAGAAAATCAATGGCACTTGGTACTACTTTGACA
GTTCAGGCTATATGCTTGCAGACCGCTGGAGGAAGCACACAGACGGCAACTGGTACTGGTTCGACAACTCAGG
CGAAATGGCTACAGGCTGGAAGAAAATCGCTGATAAGTGGTACTATTTCAACGAAGAAGGTGCCATGAAGACA
GGCTGGGTCAAGTACAAGGACACTTGGTACTACTTAGACGCTAAAGAAGGCGCCatgcaatacatcaaggcta
actctaagttcattggtatcactgaaggcgtcATGGTATCAAATGCCTTTATCCAGTCAGCGGACGGAACAGG
CTGGTACTACCTCAAACCAGACGGAACACTGGCAGACAGGCCAGAA

ATGgcGGCCAGATTTCCTTCAATTTTTACTGCAGTTTTATTCGCAGCATCCTCCGCATTAGCggccgcttACG
TACATTCCGACGGCTCTTATCCAAAAGACAAGTTTGAGAAAATCAATGGCACTTGGTACTACTTTGACAGTTC
AGGCTATATGCTTGCAGACCGCTGGAGGAAGCACACAGACGGCAACTGGTACTGGTTCGACAACTCAGGCGAA
ATGGCTACAGGCTGGAAGAAAATCGCTGATAAGTGGTACTATTTCAACGAAGAAGGTGCCATGAAGACAGGCT
GGGTCAAGTACAAGGACACTTGGTACTACTTAGACGCTAAAGAAGGCGCCatgcaatacatcaaggctaactc
taagttcattggtatcactgaaggcgtcATGGTATCAAATGCCTTTATCCAGTCAGCGGACGGAACAGGCTGG
TACTACCTCAAACCAGACGGAACACTGGCAGACAGGCCAGAAgctggtattacttacgttccaccattgttgt
tggaagttggtgttgaagaaaagttcatgtaCatgGTGCTGGGCATTGGTCCAGTGCTGGGCCTGGTCTGTGT
CCCGCTCCTAGGCTCAGCCAGTGACCACTGGCGTGGACGCTATGGCCGCCGCCGGCCCTTCATCTGGGCACTG
TCCTTGGGCATCCTGCTGAGCCTCTTTCTCATCCCAAGGGCCGGCTGGCTAGCAGGGCTGCTGTGCCCGGATC
CCAGGCCCCTGGAGCTGGCACTGCTCATCCTGGGCGTGGGGCTGCTGGACTTCTGTGGCCAGGTGTGCTTCAC
TCCACTGGAGGCCCTGCTCTCTGACCTCTTCCGGGACCCGGACCACTGTCGCCAGGCCTACTCTGTCTATGCT
TCATGATCAGTCTTGGGGGCTGCCTGGGCTACCTCCTGCCTGCCATTGACTGGGACACCAGTGCCCTGGCCCC
CTACCTGGGCACCCAGGAGGAGTGCCTCTTTGGCCTGCTCACCCTCATCTTCCTCACCTGCGTAGCAGCCACA
CTGCTGGTGGCTGAGGAGGCAGCGCTGGGCCCCACCGAGCCAGCAGAAGGGCTGTCGGCCCCCTCCTTGTCGC
CCCACTGCTGTCCATGCCGGGCCCGCTTGGCTTTCCGGAACCTGGGCGCCCTGCTTCCCCGGCTGCACCAGCT
GTGCTGCCGCATGCCCCGCACCCTGCGCCGGCTCTTCGTGGCTGAGCTGTGCAGCTGGATGGCACTCATGACC
TTCACGCTGTTTTACACGGATTTCGTGGGCGAGGGGCTGTACCAGGGCGTGCCCAGAGCTGAGCCGGGCACCG
AGGCCCGGAGACACTATGATGAAGGCGTTCGGATGGGCAGCCTGGGGCTGTTCCTGCAGTGCGCCATCTCCCT
GGTCTTCTCTCTGGTCATGGACCGGCTGGTGCAGCGATTCGGCACTCGAGCAGTCTATTTGGCCAGTGTGGCA
GCTTTCCCTGTGGCTGCCGGTGCCACATGCCTGTCCCACAGTGTGGCCGTGGTGACAGCTTCAGCCGCCCTCA
CCGGGTTCACCTTCTCAGCCCTGCAGATCCTGCCCTACACACTGGCCTCCCTCTACCACCGGGAGAAGCAGGT
GTTCCTGCCCAAATACCGAGGGGACACTGGAGGTGCTAGCAGTGAGGACAGCCTGATGACCAGCTTCCTGCCA

GGCCCTAAGCCTGGAGCTCCCTTCCCTAATGGACACGTGGGTGCTGGAGGCAGTGGCCTGCTCCCACCTCCAC
CCGCGCTCTGCGGGGCCTCTGCCTGTGAtGTCTCCGTACGTGTGGTGGTGGGTGAGCCCACCGAGGCCAGGGT
GGTTCCGGGCCGGGGCATCTGCCTGGACCTCGCCATCCTGGATAGTGCCTTCCTGCTGTCCCAGGTGGCCCCA
TCCCTGTTTATGGGCTCCATTGTCCAGCTCAGCCAGTCTGTCACTGCCTATATGGTGTCTGCCGCAGGCCTGG
GTCTGGTCGCCATTTACTTTGCTACACAGGTAGTATTTGACAAGAGCGACTTGGCCAAATACTCAGCGggtgg
acaccatcaccatcaccattaa

Construct 5 - Coding sequence of alphaprepro-P501$_{51-553}$ HIS ((in plasmid pRIT 15068 and yeast strain Y1790)

Protein sequence

MSFLNFTAVL FAASSALAAP VNTTTEDETA QIPAEAVIGY SDLEGDFDVA VLPFSNSTNN 60
GLLFINTTIA SIAAKEEGVS LEKREAEAMV LGIGPVLGLV CVPLLGSASD HWRGRYGRRR 120
PFIWALSLGI LLSLFLIPRA GWLAGLLCPD PRPLELALLI LGVGLLDFCG QVCFTPLEAL 180
LSDLFRDPDH CRQAYSVYAF MISLGGCLGY LLPAIDWDTS ALAPYLGTQE ECLFGLLTLI 240
FLTCVAATLL VAEEAALGPT EPAEGLSAPS LSPHCCPCRA RLAFRNLGAL LPRLHQLCCR 300
MPRTLRRLFV AELCSWMALM TFTLFYTDFV GEGLYQGVPR AEPGTEARRH YDEGVRMGSL 360
GLFLQCAISL VFSLVMDRLV QRFGTRAVYL ASVAAFPVAA GATCLSHSVA VVTASAALTG 420
FTFSALQILP YTLASLYHRE KQVFLPKYRG DTGGASSEDS LMTSFLPGPK PGAPFPNGHV 480
GAGGSGLLPP PPALCGASAC DVSVRVVVGE PTEARVVPGR GICLDLAILD SAFLLSQVAP 540
SLFMGSIVQL SQSVTAYMVS AAGLGLVAIY FATQVVFDKS DLAKYSAGGH HHHHH 595

Nucleotide sequence

ATGAGTTTCC TCAATTTTAC TGCAGTTTTA TTCGCAGCAT CCTCCGCATT AGCTGCTCCA 60
GTCAACACTA CAACAGAAGA TGAAACGGCA CAAATTCCGG CTGAAGCTGT CATCGGTTAC 120
TCAGATTTAG AAGGGGATTT CGATGTTGCT GTTTTGCCAT TTTCCAACAG CACAAATAAC 180
GGGTTATTGT TTATAAATAC TACTATTGCC AGCATTGCTG CTAAAGAAGA AGGGGTATCT 240
CTCGAGAAAA GAGAGGCTGA AGCCatgGTG CTGGGCATTG GTCCAGTGCT GGGCCTGGTC 300
TGTGTCCCGC TCCTAGGCTC AGCCAGTGAC CACTGGCGTG GACGCTATGG CCGCCGCCGG 360
CCCTTCATCT GGGCACTGTC CTTGGGCATC CTGCTGAGCC TCTTTCTCAT CCCAAGGGCC 420
GGCTGGCTAG CAGGGCTGCT GTGCCCGGAT CCCAGGCCCC TGGAGCTGGC ACTGCTCATC 480
CTGGGCGTGG GGCTGCTGGA CTTCTGTGGC CAGGTGTGCT TCACTCCACT GGAGGCCCTG 540
CTCTCTGACC TCTTCCGGGA CCCGGACCAC TGTCGCCAGG CCTACTCTGT CTATGCCTTC 600
ATGATCAGTC TTGGGGGCTG CCTGGGCTAC CTCCTGCCTG CCATTGACTG GGACACCAGT 660
GCCCTGGCCC CCTACCTGGG CACCCAGGAG GAGTGCCTCT TTGGCCTGCT CACCCTCATC 720
TTCCTCACCT GCGTAGCAGC CACACTGCTG GTGGCTGAGG AGGCAGCGCT GGGCCCCACC 780
GAGCCAGCAG AAGGGCTGTC GGCCCCCTCC TTGTCGCCCC ACTGCTGTCC ATGCCGGGCC 840
CGCTTGGCTT TCCGGAACCT GGGCGCCCTG CTTCCCCGGC TGCACCAGCT GTGCTGCCGC 900
ATGCCCCGCA CCCTGCGCCG GCTCTTCGTG GCTGAGCTGT GCAGCTGGAT GGCACTCATG 960
ACCTTCACGC TGTTTTACAC GGATTTCGTG GGCGAGGGGC TGTACCAGGG CGTGCCCAGA 1020

GCTGAGCCGG GCACCGAGGC CCGGAGACAC TATGATGAAG GCGTTCGGAT GGGCAGCCTG 1080
GGGCTGTTCC TGCAGTGCGC CATCTCCCTG GTCTTCTCTC TGGTCATGGA CCGGCTGGTG 1140
CAGCGATTCG GCACTCGAGC AGTCTATTTG GCCAGTGTGG CAGCTTTCCC TGTGGCTGCC 1200
GGTGCCACAT GCCTGTCCCA CAGTGTGGCC GTGGTGACAG CTTCAGCCGC CCTCACCGGG 1260
TTCACCTTCT CAGCCCTGCA GATCCTGCCC TACACACTGG CCTCCCTCTA CCACCGGGAG 1320
AAGCAGGTGT TCCTGCCCAA ATACCGAGGG GACACTGGAG GTGCTAGCAG TGAGGACAGC 1380
CTGATGACCA GCTTCCTGCC AGGCCCTAAG CCTGGAGCTC CCTTCCCTAA TGGACACGTG 1440
GGTGCTGGAG GCAGTGGCCT GCTCCCACCT CCACCCGCGC TCTGCGGGGC CTCTGCCTGT 1500
GAtGTCTCCG TACGTGTGGT GGTGGGTGAG CCCACCGAGG CCAGGGTGGT TCCGGGCCGG 1560
GGCATCTGCC TGGACCTCGC CATCCTGGAT AGTGCCTTCC TGCTGTCCCA GGTGGCCCCA 1620
TCCCTGTTTA TGGGCTCCAT TGTCCAGCTC AGCCAGTCTG TCACTGCCTA TATGGTGTCT 1680
GCCGCAGGCC TGGGTCTGGT CGCCATTTAC TTTGCTACAC AGGTAGTATT TGACAAGAGC 1740
GACTTGGCCA AATACTCAGC Gggtggacac catcaccatc accattaa 1788

Figure 8A.

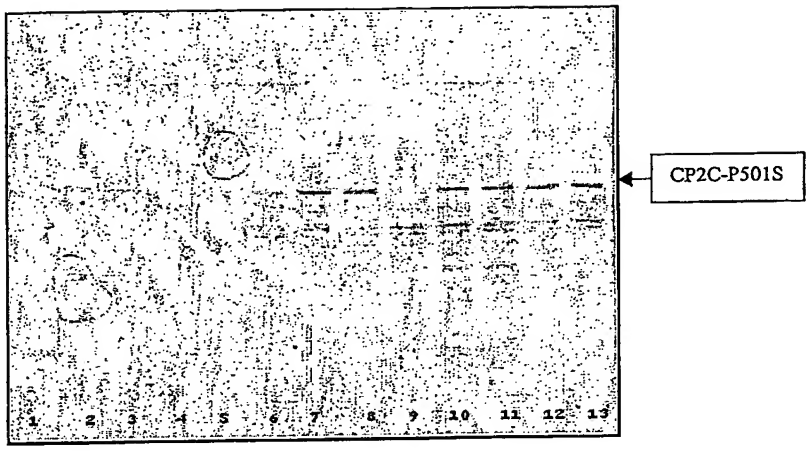
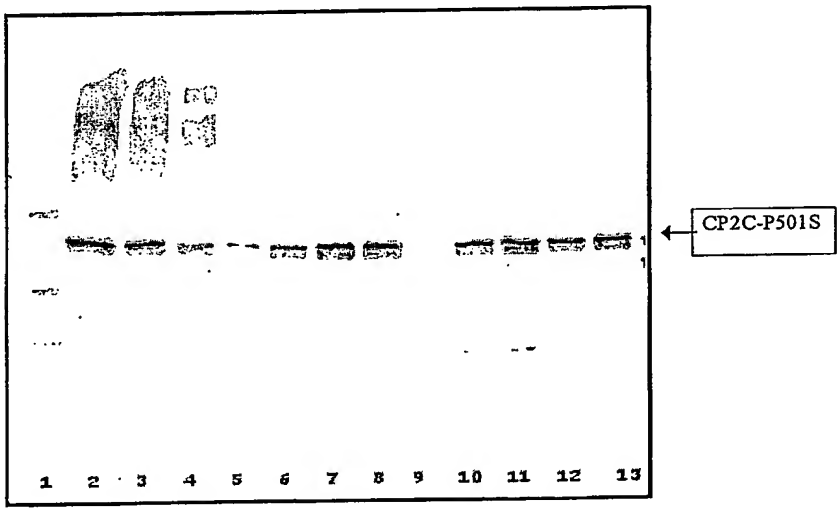


Figure 8B.



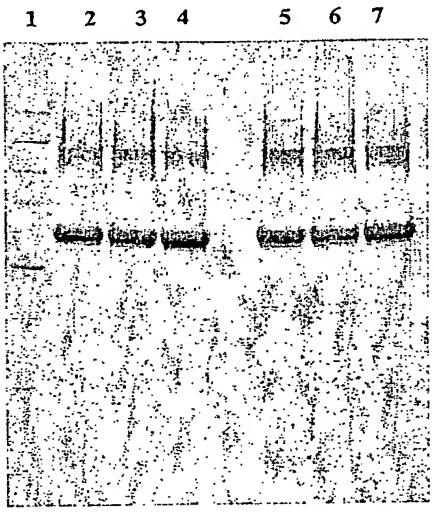
1 - **Molecular Weight Marker (Biolabs - Grow Range)**175; 83; 62; 47.5; 32.5; 25; 16.5; 6.5 kD - 10
2 - Purified Reference CP2CP501S/12 135 ng
3 - Purified Reference CP2CP501S/12 67.8 ng
4 - Purified Reference CP2CP501S/12 33.9 ng
5 - Purified Reference CP2CP501S/12 16.9 ng
6 - Fermentation PRO119-21h30
7 - Fermentation PRO124-21h30
8 - Fermentation PRO124-22h30
9 - Fermentation PRO127-0 h
10 - Fermentation PRO127-4 h
11 - Fermentation PRO127-6 h
12 - Fermentation PRO127-22h20
13 - Fermentation PRO127-22h45

Figure 9. Purification of CPC-P501-His produced by Y1796.

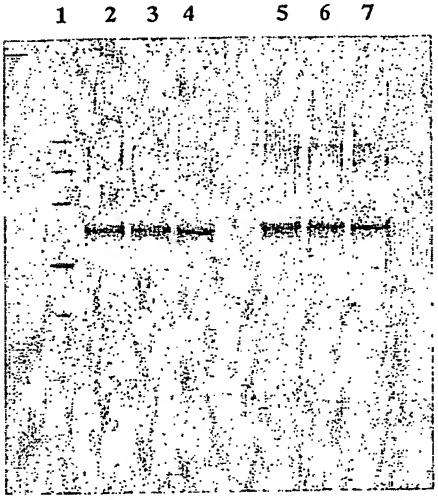
S. Cerevisiae cells	
↓	
Dyno-mill disruption	OD 120 / 2 passes / 20 mM Tris pH 8.5 - 5 mM EDTA
↓	
Centrifugation	12.000 g / RT / 90 min (supernatant discarded)
↓	
Pellet washing step 1	20 mM Tris pH 8.5 - 0.15 M NaCl - 2.0 M Guanidine.HCl - 0.1% Empigen (30 min / RT)
↓	
Centrifugation	12.000 g / RT / 60 min (supernatant discarded)
↓	
Pellet washing step 2	20 mM Tris pH 8.5 - 0.15 M NaCl - 4.0 M Urea
↓	
Centrifugation	12.000 g / RT / 30 min (supernatant discarded)
↓	
Solubilisation / Reduction	20 mM Tris pH 8.5 - 0.15 M NaCl - 8.0 M Urea - 1% SDS - 0.2 M Glutathion (60 min / RT)
↓	
Centrifugation	12.000 g / RT / 30 min (pellet discarded)
↓	
Carbamidomethylation	0.3 M Iodoacetamide (30 min / RT / in the dark) / pH adjusted to 8.5 (with 5 M NaOH solution) before incubation
↓	
R/C Supernatant	
↓	
10-fold dilution and pH adjustment (8.5)	Dilution buffer: 20 mM Tris pH 8.5 - 1 M NaCl - 8.0 M Urea
↓	
Immobilised metal ion affinity chromatography on Ni^{++}-Chelating Sepharose FF (Amersham) (10x25 cm column – 2000 ml)	Equilibration buffer: 20 mM Tris pH 8.5 - 0.9 M NaCl - 8.0 M Urea - 0.1% SDS Washing buffers: 1) Equilibration buffer 2) 20 mM Tris pH 8.5 - 0.15 M NaCl - 8.0 M Urea -

	0.1% SDS 3) 20 mM Tris pH 8.5 - 8.0 M Urea - 0.1% Tween 80 Elution buffer: 20 mM Tris pH 8.5 - 8.0 M Urea - 0.1% Tween 80 - 0.5 M Imidazole
↓	
2-fold dilution and pH adjustment (10.0)	20 mM Piperazine pH 10.0 - 8.0 M Urea - 0.1% Tween 80
↓	
Anion exchange chromatography on Q Sepharose FF (Amersham) (2,6 x 6.5 cm column - 35 ml)	Equilibration buffer: 20 mM Piperazine pH 10.0 - 8.0 M Urea - 0.1% Tween 80 Washing buffers: 1) Equilibration buffer 2) 20 mM Tris pH 8.5 - 8.0 M Urea - 0.1% Tween 80 Elution buffer: 20 mM Tris pH 7.5 - 8.0 M Urea - 0.1% Tween 80 - 0.5 M NaCl
↓	
Concentration/Diafiltration (Pall - Omega 10 kDa - 200 cm²)	+/- 3-fold concentration Diafiltration buffer: Tris 20 mM pH 7.5
↓	
Sterile filtration (Millipore - Millex GV 0.22µm)	
↓	
Purified bulk	Final buffer: 20 mM Tris pH 7.5 - +/- 0.3% Tween 80
↓	
Storage –20°C	

Figure 10. Pattern of CPC P501 His purified protein (4-12% Novex Nu-Page polyacrylamide precasted gels)



Coomassie Blue R250



Daiichi Silver Staining

Western Blot anti P501S (Monoclonal antibody)

1: MW (250/150/75/50/37/25/15/10 kDa)
2: Purified bulk A (reducing conditions)
3: Purified bulk B (reducing conditions)
4: Purified bulk C (reducing conditions)
5: Purified bulk A (non reducing conditions)
6: Purified bulk B (non reducing conditions)
7: Purified bulk C (non reducing conditions)